

De techniek van gebruikersonderzoek

HENK VOORBIJ

We leven in een gebruikersgericht klimaat. We doen ons werk voor de klant. We proberen onze producten, diensten en voorzieningen zo goed mogelijk af te stemmen op de behoeften van de klant. Het ligt dan ook voor de hand dat we er veel waarde aan hechten de stem van de klant te horen. Een (periodiek) gebruikersonderzoek of klanttevredenheidsonderzoek is daartoe een geëigende methode. Dit hoofdstuk is bedoeld als een praktische handleiding voor alle activiteiten die daarbij een rol spelen: ontwerp van een vragenlijst, samenstelling van een steekproef en analyse van de gegevens.

Deze bijdrage is als volgt opgebouwd: Paragraaf 1 zet de term gebruikersonderzoek af tegen verwante begrippen als gebruiksonderzoek en survey. Paragraaf 2 laat zien dat op het gebied van informatiewetenschap veelvuldig gebruikersonderzoek wordt toegepast. De paragrafen 3 en 4 gaan in op het ontwerp van de vragenlijst: hoe bedenkt je relevante vragen en hoe formuleer je deze. Paragraaf 5 stelt de ontwikkeling van een bijzonder soort vragenlijsten aan de orde: een schaal met als uitkomst een totaalscore. Wanneer de vragenlijst gereed is, dient deze verspreid te worden. Paragraaf 6 behandelt de verschillende mogelijkheden. De wie-vraag komt aan bod in paragraaf 7. Gebruikersonderzoek richt zich meestal niet op de gehele populatie, maar op een steekproef daaruit. Hoe stellen we die samen? In paragraaf 8 zijn we beland in het stadium van de analyse van de resultaten. Deze bevat onder meer een beknopt overzicht van aspecten uit de beschrijvende statistiek. Paragraaf 9 is gewijd aan een specifieke vorm van gebruikersonderzoek: het klanttevredenheidsonderzoek. Verschillende mogelijkheden om dat vorm te geven passeren de revue. Paragraaf 10 ten slotte, zet een aantal beperkingen op een rijtje. Niet de minste daarvan is het subjectieve karakter van gebruikersonderzoek: in hoeverre zijn de antwoorden betrouwbaar?

1 Definitie

Om te achterhalen hoeveel boeken er vandaag uitgeleend zijn, kunnen we twee methoden toepassen: raadplegen van de uitleenadministratie, of de hele dag gebruikers bij het verlaten van de bibliotheek ondervragen. In het eerste geval doen we een gebruiksonderzoek (use study), in het tweede een gebruikersonderzoek (user study). Er is sprake van gebruikersonderzoek als we rechtstreeks een beroep doen op de gebruiker om onze gegevens te verzamelen. Daarbij denken we aan vragenlijsten, interviews en observatie. De term gebruikersonderzoek duidt dus niet op het onderwerp van het onderzoek (wat), maar op de methode om gegevens te verzamelen (hoe).

Vragenlijsten worden vaak gebruikt bij surveys. Survey betekent overzicht. Kenmerken zijn:

- Men wil de huidige stand van zaken op een bepaald onderwerp in kaart brengen.
- Het onderzoek heeft betrekking op een relatief kleine steekproef; de resultaten zouden representatief moeten zijn voor de gehele populatie.
- De gegevens worden verzameld door middel van vragenlijsten, interviews of observatie.

De OECD hanteert de volgende definitie:

‘A survey is an investigation about the characteristics of a given population by means of collecting data from a sample of that population and estimating their characteristics through the systematic use of statistical methodology.’¹

In de praktijk worden de begrippen survey en gebruikersonderzoek soms als identiek beschouwd. Dit is niet helemaal terecht. Een gebruikersonderzoek kan zich ook richten op een complete, duidelijk afgegrensde populatie. Anderzijds kan een survey betrekking hebben op objecten, in plaats van mensen. Denk bijvoorbeeld aan een onderzoek naar de verzuringsgraad van boeken aan de hand van een steekproef uit de collectie. In veel gevallen echter komen survey en gebruikersonderzoek op hetzelfde neer. Wie wil weten wat de ervaringen van studenten zijn met

Google Scholar, zal er vaak voor kiezen een vragenlijst te verspreiden onder een steekproef van de studentenpopulatie.

2 Belang

Gebruikersonderzoek is een van de meest toegepaste methoden op het gebied van de informatiewetenschap. Hider en Pymm (2008) onderzochten 834 artikelen, opgenomen in jaargang 2005 van twintig vooraanstaande tijdschriften. Daarvan konden er 567 (68,0%) worden getypeerd als research artikel. De meest voorkomende onderzoeksmethode was de survey (30,5%). Experiment was een goede tweede, met 20,8%. Andere, minder toegepaste methoden waren bijvoorbeeld case study, kwalitatief onderzoek, inhoudsanalyse en citaatanalyse. Gezien het overwicht van survey als methode wekt het geen verbazing dat vragenlijsten en interviews de belangrijkste techniek waren om gegevens te verzamelen (32,5%).

Onder de twintig onderzochte tijdschriften bevinden zich titels als *Management Information Systems Quarterly*, *Information Systems Research*, *International Journal of Geographical Information Science* en *Research Evaluation*. Zes titels bevinden zich op het terrein van bibliotheekwetenschap, waaronder *College & Research Libraries*, *Journal of Academic Librarianship* en *Library & Information Science Research*. Afgaande op alleen die zes titels is het aandeel van surveys met 51,4% en vragenlijsten en interviews met 42,9% nog een stuk hoger. Opvallend is dat experimenten op het terrein van bibliotheekwetenschap geen enkele rol spelen.

Andere studies bevestigen het beeld. Julien, Pecoskie en Reed (2011) onderzochten 749 artikelen op het meer specifieke gebied van informatiegedrag, gepubliceerd tussen 1999 en 2008. In 528 van deze artikelen (70,6%) werd verslag gedaan van onderzoek. De daarbij gehanteerde methoden waren vragenlijsten (34,7%, n=183), interviews (10,0%, n=53), transactie log analyse (9,8%, n=52), experiment (5,3%, n=28), citaatanalyse (1,9%, n=10) en etnografie (1,3%, n=7). Turcios, Agarwal en Watkins (2014) richtten zich niet op de toptijdschriften, maar op alle 105 tijdschriften op het gebied van bibliotheek- en informatiewetenschap waarop hun bibliotheek een abonnement had. Zij bestudeerden in totaal 1880 artikelen uit de meest recente afleveringen.

Omdat een groot aantal tijdschriften vooral een praktische oriëntatie heeft, is het aandeel van researchartikelen een stuk lager (16,3%, n=307). Binnen die categorie was de survey met 21% de meest toegepaste methode. Tuomaala, Järvelin en Vakkari (2014) tenslotte, hanteerden een methodiek die sterk verwant is aan die van Hider en Pymm. Het bijzondere van hun studie is dat zij een vergelijking maakten tussen 1965, 1985 en 2005. In alle drie perioden staat de survey aan kop, met een aandeel van 22 tot 25%. Het aandeel van het experiment nam toe van nog geen 10% in de eerste twee perioden naar 22% in de laatste periode.

3 Vragen bedenken

Gebruikersonderzoek kan zich richten op tal van vraagstukken:

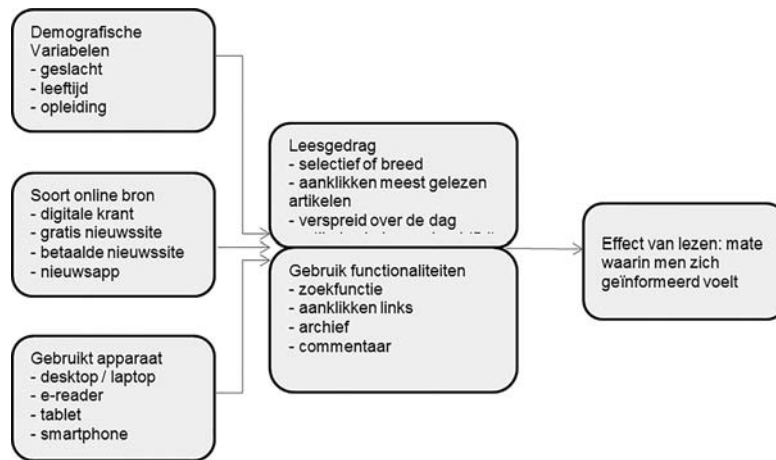
- het gebruik van online nieuwsbronnen door studenten;
- de ervaringen van bezoekers met de website Delpher;
- tevredenheid over de dienstverlening van de bibliotheek;
- de omgang met wetenschappelijke artikelen door wetenschappers (King et al, 2009);
- de bereidheid om eigen publicaties in te voeren in een repository (Kim, 2011).

Geen enkele onderzoeker kan zomaar over al deze onderwerpen een vragenlijst uit zijn mouw schudden. Een eerste vereiste is dat je je verdiept in het onderwerp en weet welke aspecten allemaal een rol spelen. Pas dan kun je dat vertalen in relevante vragen met duidelijke antwoordmogelijkheden. Deze kennis krijg je door de literatuur te bestuderen of gesprekken te voeren met de opdrachtgever of betrokkenen.

Laten we als voorbeeld de vraag naar het gebruik van online nieuwsbronnen door studenten nemen. Welke soorten online nieuwsbronnen zijn er eigenlijk? Op welke apparaten kunnen deze geraadpleegd worden? Welke mogelijkheden bieden deze bronnen? En hoe kunnen we het leesgedrag typeren? Verschilt dit per type bron? En hangt de omgang met online nieuwsbronnen samen met demografische variabelen zoals leeftijd, geslacht en opleidingsniveau?

Deze vragen zijn schematisch weergegeven in figuur 1. Deze bevat niet alleen een meer concrete invulling van de aspecten, maar brengt ook de

veronderstelde relaties in beeld. We spreken daarom van een conceptueel model. Links staan de onafhankelijke variabelen, rechts de afhankelijke variabelen. We veronderstellen dat de variabelen aan de linkerkant van invloed zijn op de variabelen aan de rechterkant, en geven dat weer met een pijltje.



Figuur 1. Conceptueel model omgang met online nieuwsbronnen.

Een conceptueel model stuurt het onderzoek en is daarom van wezenlijk belang. Het geeft in één oogopslag weer welke variabelen een rol spelen en wat het mogelijke verband daartussen is. En daarmee welke vragen we moeten stellen en welke (kruis)tabellen we moeten vervaardigen. Bijvoorbeeld een kruistabel die de relatie in beeld brengt tussen het aantal raadplegingen per dag en het type bron.

| | Digitale krant | Gratis nieuwssite | Betaalde nieuwssite | Nieuwsapp |
|----------------|-----------------------|--------------------------|----------------------------|------------------|
| 1 × per dag | | | | |
| 2-3 × per dag | | | | |
| 4-10 × per dag | | | | |
| > 10 × per dag | | | | |

4 Vragen formuleren

4.1 Woordkeuze

Om betrouwbare antwoorden te krijgen moeten vragen ondubbelzinnig zijn: elke respondent moet de vraag begrijpen en op dezelfde manier opvatten. Dit lijkt gemakkelijker dan het is. Er liggen veel valkuilen op de loer. Deze paragraaf stelt er een aantal aan de orde.

1. Terminologie, jargon

Voor informatieprofessionals zijn termen als database, functionaliteit, interface, OCR en deep linking gemeengoed. Het is maar de vraag of dat ook opgaat voor de respondenten. Wellicht denken zij bij databases aan een Excel spreadsheet. Het is daarom raadzaam om in dit soort gevallen een kleine omschrijving of toelichting toe te voegen. Dit is lang niet altijd eenvoudig. Hoe maak je de essentie van OCR in een zin duidelijk?

Wat zou prioriteit moeten hebben bij de verdere ontwikkeling van Delpher?

- a. Meer kranten digitaliseren.
- b. Optical Character Recognition (OCR) verbeteren (de omzetting van scans in voor de computer herkenbare tekens, zodat de pagina's met een bepaalde zoekterm ook daadwerkelijk gevonden kunnen worden).

2. (Dubbele) ontkenningen

| | 1. Helemaal oneens | 2. Oneens | 3. Niet eens, niet oneens | 4. Mee eens | 5. Helemaal mee eens |
|--|--------------------|-----------|---------------------------|-------------|----------------------|
| a. Illegaal downloaden van muziek is helemaal niet slecht voor een artiest | | | | | |
| b. Ik ben niet tevreden over het aanbod op Spotify | | | | | |

Wie vindt dat illegaal downloaden van muziek wel slecht is voor een artiest, moet hier schaalwaarde 1 of 2 kiezen. Hij is het er (helemaal) mee oneens dat illegaal downloaden niet slecht is voor een artiest. Het is denkbaar dat sommige respondenten door deze dubbele ontkenning in de war raken. Voor de tweede bewering geldt hetzelfde.

3. *Te voorzichtige stellingen*

Stellingen moeten krachtig geformuleerd worden. De derde optie ('Ik ben erg tevreden over...') verdient dan ook de voorkeur. Bij de eerste en tweede bewering komen mensen in de knoop die zeer tevreden zijn: als ik heel tevreden ben, kan ik toch niet zeggen dat ik het helemaal eens ben met de kwalificatie redelijk tevreden?

| | 1. Helemaal oneens | 2. Oneens | 3. Niet eens, niet oneens | 4. Mee eens | 5. Helemaal mee eens |
|--|--------------------|-----------|---------------------------|-------------|----------------------|
| a. Ik ben redelijk tevreden over het aanbod op Spotify | | | | | |
| b. Ik ben tevreden over het aanbod op Spotify | | | | | |
| c. Ik ben erg tevreden over het aanbod op Spotify | | | | | |

4. *Suggestieve vragen*

Koopt u nog cd's?
Zo ja: hoeveel cd's heeft u het afgelopen jaar gekocht?

Dit is een voorbeeld van een suggestieve vraag. Kennelijk ben je een beetje ouderwets als je (nog!) cd's koopt. Het is denkbaar dat men een lager aantal invult dan in werkelijkheid het geval is, om niet al te stoffig over te komen.

5. *Gevoelswaarde van termen*

In 1941 werd een experiment gedaan om de invloed van woordkeuze te onderzoeken. De ene helft van de respondenten kreeg een vraag voorgelegd met daarin de term ‘verbieden’, de andere helft met ‘niet toestaan’.

- a. Do you think the United States should forbid public speeches against democracy?
- b. Do you think the United States should not allow public speeches against democracy?

Met het eerste alternatief was 54% het eens, met het tweede 75%. Aan ‘verbieden’ kleefte kennelijk een andere gevoelswaarde dan aan ‘niet toestaan’. Het experiment is daarna nog verschillende keren herhaald, met wisselend resultaat.²

Het experiment toont aan dat onderzoekers, al dan niet bewust, de resultaten kunnen manipuleren door de exacte formulering van de vraag. Temeer daar bij het grote publiek vooral de getallen blijven hangen. Men zal zich herinneren dat 75% (of 54%) tegen anti-democratische toespraken is, zonder te weten hoe de vraag precies geformuleerd was.

6. *Positieve of negatieve formulering van een stelling*

Er is eveneens een verschil in gevoelswaarde tussen een positieve en een negatieve uitspraak. Iemand die schaalwaarde 1 toekent op de eerste vraag, zou schaalwaarde 5 moeten toekennen op de tweede vraag. Het is echter maar de vraag of dit in werkelijkheid gebeurt. Mensen vinden het vaak moeilijker om iets ‘slecht’ te noemen dan ‘niet goed’. Mogelijk beantwoordt men de tweede vraag niet met schaalwaarde 5, maar 4. De kans op een dergelijk effect is nog groter bij een zeven- of negenpuntschaal: de uitwijkmogelijkheden zijn dan groter.

| | 1. Helemaal oneens | 2. Oneens | 3. Niet eens, niet oneens | 4. Mee eens | 5. Helemaal mee eens |
|--|--------------------|-----------|---------------------------|-------------|----------------------|
| a. De website is heel overzichtelijk | | | | | |
| b. De website is heel onoverzichtelijk | | | | | |

4.2 Vraagtypen

Er is een sterke voorkeur om bij vragenlijsten te werken met gesloten vragen: vragen met voorgeformuleerde antwoordmogelijkheden. Open vragen kunnen wel veel informatie opleveren, maar zijn bewerkelijk voor de respondent om te beantwoorden en voor de onderzoeker om te verwerken. Toch is het vaak nuttig een beperkt aantal open vragen op te nemen:

- Ter nadere toelichting, als vervolg op een eerder gestelde vraag.
- Ter verkenning, op aspecten die nog niet goed in kaart zijn gebracht, of waar de onderzoeker open staat voor verrassende antwoorden: ‘Wat bevalt u het meest aan onze website?’ of ‘In welke opzichten komt onze website in aanmerking voor verbetering?’
- Ter uitsluiting van het al te gemakkelijk aankruisen van antwoordmogelijkheden. Margaret Stieg Dalton (2004) onderzocht in 1981 en 2003 van welke bestanden historici gebruik maken. In het eerste onderzoek legde zij respondenten een kant-en-klaar lijstje van titels voor. Veel historici bleken titels aan te kruisen waarvan zij vonden dat zij die zouden moeten gebruiken of waarvan ze alleen de naam kenden. In het tweede onderzoek dienden respondenten spontaan titels te noemen. Het is waarschijnlijk dat zij de opgevoerde bestanden ook daadwerkelijk gebruiken.

Een verkeerde reden om open vragen te stellen is dat de onderzoeker onvoldoende kennis heeft van het onderwerp. Een onderzoeker die zich niet verdiept in redenen waarom iemand geen online nieuwsbronnen gebruikt, kan zich er gemakkelijk van afmaken door kortweg te vragen: ‘Waarom leest u geen nieuws in digitale vorm?’ Een gesloten vraag met vooraf geformuleerde antwoordmogelijkheden is hier zeer goed mogelijk. De respondenten kunnen volstaan met het aanvinken van de opties die van toepassing zijn, of met het invullen van een schaalwaarde.

Waarom gebruikt u geen online nieuwsbronnen (meerdere antwoorden mogelijk)?

- a. Ik ben gehecht aan de meer traditionele media, zoals papieren kranten, TV, radio
- b. Ik vind het niet belangrijk om overal en 24 uur per dag het nieuws te kunnen volgen
- c. Etcetera

Waarom gebruikt u geen online nieuwsbronnen? Geef aan in hoeverre u het eens bent met de volgende stellingen

| | Helemaal oneens | Oneens | Niet eens, niet oneens | Mee eens | Helemaal mee eens |
|--|-----------------|--------|------------------------|----------|-------------------|
| a. Ik ben gehecht aan de meer traditionele media, zoals papieren kranten, TV, radio | | | | | |
| b. Ik vind het niet belangrijk om overal en 24 uur per dag het nieuws te kunnen volgen | | | | | |
| c. Etcetera | | | | | |

4.3 Antwoordcategorieën

1. Volledigheid

Probeer bij vragen met vooraf geformuleerde antwoordmogelijkheden, zoals de redenen van gebruik of niet-gebruik van een bepaalde dienst, zo volledig mogelijk te zijn. Je kunt echter niet van tevoren alles bedenken. Sluit het rijtje daarom af met: 'overige, namelijk...' Hopelijk komt maar een klein deel van de antwoorden in deze categorie terecht. Zo niet, dan moeten we voorzichtig zijn met conclusies. We hebben kennelijk niet de meest relevante zaken voorgelegd.

2. *Geen vage termen*

Vermijd antwoordmogelijkheden zoals soms, regelmatig en vaak. Wat voor de één soms is, is voor de ander vaak. Koppel een vraag naar de frequentie van een activiteit bovendien aan een periode: de afgelopen week, de laatste drie maanden... De formulering 'de afgelopen twaalf maanden' is beter dan 'het afgelopen jaar', omdat in het laatste geval de neiging bestaat te denken aan een kalenderjaar.

Hoeveel keer heeft u Delpher bezocht in de afgelopen twaalf maanden?

- 0 keer
- 1-3 keer
- 4-10 keer
- 11-30 keer
- Meer dan 30 keer

3. *Exacte aantallen of categorieën*

Je kunt niet verwachten dat iemand exact weet hoe vaak hij in een afgelopen periode een website heeft bezocht of een boek heeft gelezen. Daarom leggen we categorieën voor waar iedereen mee uit de voeten kan. Bij leeftijd hebben we de keus tussen exacte leeftijd of leeftijdscategorie. Een voordeel van exacte leeftijden is dat we bij de analyse alle mogelijke leeftijdscategorieën kunnen samenstellen.

4. *Aantal schaalwaarden*

Hoe tevreden bent u over de openingstijden van de bibliotheek? Hieronder zijn enkele opties afgebeeld.

| Optie A | Optie B | Optie C | Optie D | Optie E |
|---------------|--------------------|--------------------|------------------|--------------------|
| 1. Ontevreden | 1. Zeer ontevreden | 1. Zeer ontevreden | 1. Ontevreden | 1. Zeer ontevreden |
| 2. Neutraal | 2. Ontevreden | den | 2. Enigszins | 2. |
| 3. Tevreden | 3. Neutraal | 2. Ontevreden | tevreden | 3. |
| | 4. Tevreden | 3. Tevreden | 3. Tevreden | 4. |
| | 5. Zeer tevreden | 4. Zeer tevreden | 4. Zeer tevreden | 5. |
| | | | | 6. |
| | | | | 7. Zeer tevreden |

Optie A biedt mensen weinig ruimte. Bovendien zijn mensen minder geneigd een van de uitersten te kiezen, en dus naar het midden te vluchten. Zij kunnen dan maar op één plaats terecht.

Optie B is vrij gebruikelijk. Een vijfpuntsschaal is een mooie balans tussen subtiliteit en hanteerbaarheid.

De opties C en D hebben een vierpuntsschaal. Bij een even aantal antwoordcategorieën kan men mensen dwingen een keuze te maken; meer naar het tevredene, dan wel meer naar het ontevredene. Er is hierbij geen middencategorie met een neutraal antwoord. Binnen deze twee opties bestaat een lichte voorkeur voor de – weliswaar minder evenwichtige – optie D. Het verschil tussen ontevreden en zeer ontevreden is niet zo relevant, terwijl de zeer tevreden klanten de loyale klanten zijn. Deze zijn zeer belangrijk voor een organisatie.

Optie E is een zevenpuntsschaal. Hierbij zijn alleen de uitersten benoemd, de taal schiet te kort om alle tussenwaarden in woorden te vangen.

Bij een oneven aantal antwoordmogelijkheden is de middelste gewoonlijk een neutrale. Neutraal betekent niet ‘geen mening’. In sommige gevallen is het nuttig deze optie apart toe te voegen. Bij zaken waarvan je kunt aannemen dat iedereen daar wel een mening over heeft, kun je dit beter weglaten.

Alle opties in het voorbeeld lopen van negatief naar positief. Je kunt net zo goed kiezen voor de omgekeerde volgorde. Wel is het raadzaam de hele vragenlijst door dezelfde volgorde aan te houden. Om te voorkomen dat respondenten in een ‘antwoordcadans’ terechtkomen, kun je beter de vraagstelling zelf afwisselen, door af en toe een negatief geformuleerde vraag voor te leggen: ‘het aantal zitplaatsen in de bibliotheek is volstrekt onvoldoende’. Wees daar terughoudend mee: mensen doorzien dat al gauw en vinden dat irritant.

4.4 *Algemeen of ‘critical incident’*

Bij de ‘critical incident’ techniek koppel je de vragen aan een specifieke gebeurtenis. Dit kan zijn: het laatste bezoek aan de bibliotheek of website, het laatst gelezen artikel, de keer dat je bijzonder tevreden was over de

dienstverlening, of juist ontevreden. King en Tenopir passen deze techniek veelvuldig toe bij hun onderzoeken naar de omgang met wetenschappelijke artikelen. Het voordeel is niet alleen dat mensen zich vaak beter details kunnen herinneren als de vraag is opgehangen aan een specifieke gebeurtenis, maar ook zich kunnen toespitsen op dit geval. Wetenschappers lezen artikelen om allerlei redenen: in het kader van hun onderzoek, ter ondersteuning van hun onderwijs, om op de hoogte te blijven. Zij komen artikelen ook op tal van manieren op het spoor, bijvoorbeeld door zoektermen in te typen, nieuwe afleveringen van belangrijke tijdschriften langs te lopen, aan de hand van citaties of via attentering door collega's. Maar voor het laatst gelezen artikel geldt telkens maar één van die opties, bijvoorbeeld: dit specifieke artikel is ontdekt door het browsen in een nieuwe tijdschriftaflevering en is gelezen om op de hoogte te blijven van nieuwe ontwikkelingen. Daarmee is het mogelijk om tal van kruistabellen te vervaardigen. Tabel 1 laat duidelijk zien dat browsen een belangrijke ontdekkingsbron is voor current awareness doeleinden, terwijl citaties vooral een rol spelen bij het opsporen van artikelen voor onderzoeksdoeleinden. Bij algemeen gestelde vragen zou het moeilijk zijn om dergelijke uitspraken te doen.

How readers learn about articles read for research, teaching and current awareness (n = 881)

| Purpose of reading | How readers became aware of articles (%) | | | | |
|--------------------|--|-----------|-----------|---------|-------------------|
| | Browsing | Searching | Citations | Persons | Total |
| Research | 30.9 | 29.7 | 21.2 | 18.3 | 101 |
| Teaching | 44.1 | 25.3 | 12.2 | 18.3 | 99.9 |
| Current awareness | 64.7 | 14.3 | 2.5 | 18.5 | 100 |
| All | 35.7 | 26.7 | 15.8 | 17.6 | 95.8 ^a |

^aLess than 100% because 'other' purposes not included.

Tabel 1. Kruistabel identificatie artikel – doel van lezen (naar King et al, 2009).

De 'critical incident' techniek leent zich vooral voor activiteiten die zich veelvuldig afspelen en veel variatie vertonen. Voor artikelen geldt niet alleen dat je ze op verschillende manieren kunt ontdekken en om verschillende redenen kunt lezen, maar je kunt er ook meer of minder tijd aan besteden en er meer of minder nut van ondervinden.

4.5 *Belang of gebruik*

Wat is de beste vraag:

- Hoe belangrijk voor uw werk zijn wetenschappelijke tijdschriften (uit onze collectie)?
1 heel onbelangrijk – 5 heel belangrijk.
- Hoe vaak gebruikt u voor uw werk wetenschappelijke tijdschriften (uit onze collectie)?
Dagelijks, meerdere keren per week, wekelijks, maandelijks, minder dan een keer per maand, nooit.

Belangrijk en veelvuldig gebruik zullen vaak samenvallen, maar niet altijd: iemand kan veel belang hechten aan een voorziening, maar het gebruik ervan concentreren op bepaalde perioden. In dat geval zal hij eerder de optie 'maandelijks' aankruisen dan 'dagelijks', met het gevaar van verkeerde gevolgtrekkingen. Omgekeerd is het niet gezegd dat iemand die veel belang hecht aan een bepaalde voorziening deze ook intensief gebruikt. Misschien is er eerder sprake van 'option value' (men ziet iets als waardevol op grond van mogelijk toekomstig gebruik), 'existence value' (men vindt het belangrijk dat iets bestaat ook al zal men er zelf nooit gebruik van maken), of 'bequest value' (men vindt het belangrijk een bepaalde voorziening in stand te houden voor toekomstige generaties).

Misschien is de beste keuze om te vragen naar het 'belang voor uw werk, studie, hobby', om daarmee althans existence value en bequest value uit te sluiten. Belang is ook vaker toepasbaar dan gebruik. Het is gemakkelijker om aan te geven of je een voorziening als het doorklikken naar achtergrondinformatie bij nieuwsberichten belangrijk vindt, dan om in te vullen hoe vaak je dat de afgelopen periode hebt gedaan.

4.6 *Volgorde van de vragen*

Een vragenlijst moet voor de respondent gemakkelijk en interessant zijn. Vragen naar persoonlijke gegevens zijn voor hem of haar niet interessant, maar wel gemakkelijk te beantwoorden. Er is daarom net zoveel voor te zeggen om deze vooraan als achteraan te plaatsen. Als het om een groot aantal achtergrondvragen gaat, is het beter om deze aan het eind van de vragenlijst op te nemen. Anders zou het te lang duren voordat de eigenlijke vragen aan bod komen.

Respondenten ervaren het als logisch wanneer vragen over een onderwerp gaan van algemeen naar specifiek, en van feitelijk gedrag naar mening naar suggesties voor verbetering. Op die manier bied je hen de gelegenheid zich in de materie te verdiepen, zich warm te draaien. Bij interviews begin je meestal ook met enkele ijsbrekervragen. Laat iemand eerst vertellen over bijvoorbeeld het project waarmee hij bezig is; de details komen daarna als vanzelf aan bod.

Gropeer de vragen indien mogelijk in secties, duidelijk aangegeven met aparte letters (A, B, C etc.) of kleine tekstjes: ‘de volgende vragen gaan over ...’ Dit biedt structuur, niet alleen aan de onderzoeker, maar ook aan de respondent.

De vragenlijst eindigt gewoonlijk met een open vraag: ‘Heeft u verder nog suggesties of opmerkingen?’ Niet zelden levert dit nog verrassende inzichten op. Als echter al te uitbundig van deze mogelijkheid gebruik wordt gemaakt, is dat mogelijk een teken dat de vragenlijst te kort schiet.

4.7 Vooronderzoek

Het maken van een vragenlijst is een proces van vallen en opstaan. Vaak zijn enkele revisies nodig, omdat blijkt dat de vragen toch niet eenduidig zijn, aspecten over het hoofd gezien zijn, er weinig afwisseling is in vraagtypen, of de volgorde onhandig is.

Voordat de vragenlijst zijn definitieve status krijgt en verspreid wordt, is het raadzaam een beperkt vooronderzoek te doen onder enkele personen die behoren tot de doelgroep. Let daarbij op de volgende aspecten:

- Onduidelijke vragen. Verzoek de proefpersonen die vragen aan te kruisen die niet helemaal duidelijk of voor meerdere uitleg vatbaar zijn. Vraag hen na afloop dit toe te lichten.
- Lacunes. Verzoek de proefpersonen aan te geven of zij hun ei kwijt konden, het gevoel hadden dat alle relevante aspecten in de vragenlijst aan bod komen. Mogelijk ontbreken er zaken en geeft de huidige versie geen goed beeld van wat er allemaal speelt.
- Invultijd. De begeleidende tekst waarmee een vragenlijst wordt aangeboden bevat vaak informatie over de verwachte invultijd. Houd daarom bij hoe lang de proefpersonen in het vooronderzoek ermee bezig waren.

Mogelijk vinden op basis van het vooronderzoek nog enkele laatste aanpassingen plaats.

5 Schaalconstructie

We kunnen een vragenlijst beschouwen als een verzameling afzonderlijke vragen en de antwoorden aan afzonderlijke analyses onderwerpen: het percentage respondenten dat ja aangeeft op vraag 8, de gemiddelde schaalwaarde van de antwoorden op vraag 16, een verschil tussen de scores van mannen en vrouwen op vraag 20, en de correlatie tussen de antwoorden op de vragen 22 en 23. Een dergelijke analyse kan een heel genuanceerd beeld opleveren.

In sommige gevallen heeft een vragenlijst meer het karakter van een test: uit de antwoorden maken we op hoe het gesteld is met de kennis of attitude van respondenten. Bekende voorbeelden zijn schooltoetsen en tentamens. De uitslag bestaat uit een totaalcijfer, niet uit cijfers per vraag. Op die manier kunnen we ook onderzoeken in hoeverre mensen beschikken over informatievaardigheden, in hoeverre ze milieubewust zijn, of wat hun politieke oriëntatie is. Telkens berekenen we een totaalcijfer, dat we kunnen vertalen in kwalificaties als ‘beschikkend over matige informatievaardigheden’, ‘zeer sterk milieubewust’ of ‘gematigd links’.

Om te bepalen of een verzameling vragen als een schaal beschouwd kan worden, moet het zo zijn dat proefpersonen die (relatief) hoog scoren op de ene vraag ook (relatief) hoog scoren op andere vragen. Zie onderstaand voorbeeld met de scores op vier vragen volgens een zevenpuntsschaal. Er is een duidelijke correlatie zichtbaar tussen de antwoorden op vraag 1 en vraag 2. Personen die op vraag 1 relatief hoog scoren, doen dat ook op vraag 2. Eveneens is er een duidelijke correlatie tussen de antwoorden op vraag 1 en vraag 3. Personen die op vraag 1 hoog scoren, doen dat ook op vraag 3. In absolute zin liggen de scores op vraag 3 telkens een stuk een lager dan die op vraag 1, maar in relatieve zin zijn ze vergelijkbaar. Een dergelijke correlatie ontbreekt tussen vraag 1 en 4.

| | Vraag 1 | Vraag 2 | Vraag 3 | Vraag 4 |
|-------|---------|---------|---------|---------|
| Ppn 1 | 7 | 6 | 4 | 2 |
| Ppn 2 | 5 | 4 | 3 | 5 |
| Ppn 3 | 4 | 4 | 2 | 6 |
| Ppn 4 | 2 | 2 | 1 | 4 |
| Ppn 5 | 7 | 7 | 5 | 3 |
| Ppn 6 | 7 | 6 | 4 | 7 |
| Ppn 7 | 6 | 7 | 4 | 1 |
| ... | | | | |

De onderlinge correlaties tussen alle vragen kunnen in één getal worden uitgedrukt: Cronbach's alfa. Cronbach's alfa is dus een maat voor de homogeniteit van de schaal. De formule luidt:

$$\alpha = \frac{k \times r_{\text{gemiddeld}}}{1 + (k - 1) \times r_{\text{gemiddeld}}}$$

waarbij k = het aantal items

$r_{\text{gemiddeld}}$ = het gemiddelde van alle correlaties tussen de items

(voorbeeld: bij 4 vragen zijn er 6 correlaties: 1-2, 1-3, 1-4, 2-3, 2-4, 3-4)

Om van een homogene schaal te mogen spreken, moet Cronbach's alfa tussen 0,65 en 0,90 liggen. Niet te laag, maar ook weer niet al te hoog. Bij een waarde hoger dan 0,90 hebben de vragen nog nauwelijks discriminerende waarde. Het heeft bijvoorbeeld weinig zin om zowel naar leeftijd als geboortjaar te vragen. Met SPSS kun je Cronbach's alfa gemakkelijk uitrekenen.

Een tweede voorwaarde om van een schaal te mogen spreken is dat de *item-total correlatie* voor elk item groter dan of gelijk is aan 0,30. Deze maat geeft aan in hoeverre de antwoorden op een vraag correleren met de som van de antwoorden op alle andere vragen. Idealiter heeft elke vraag een hoge correlatie met de somscore. Met SPSS kun je gemakkelijk achterhalen welke vragen niet aan de eis voldoen, en wat het effect is op Cronbach's alfa als je deze weglaat. Zo kun je successievelijk minder passende vragen verwijderen en vaststellen of dit nog leidt tot

verbetering van Cronbach's alfa. Het resultaat is dat er uiteindelijk een kleiner aantal vragen resteert. Deze vormen echter wel een beter, meer samenhangend geheel.

6 Wijze van afname

Als het enigszins kan, worden vragenlijsten online afgenomen. Er zijn diverse softwarepakketten die dit mogelijk maken. De bekendste is wellicht SurveyMonkey (www.surveymonkey.com). Voorbeelden van andere pakketten zijn:

- <http://esurv.org>
- <http://www.google.com/forms/about>
- <http://www.enqueteviainternet.nl>
- <http://www.surveygold.com>
- <http://www.formdesk.nl>
- <http://www.questionpro.com>

Deze ondersteunen alle mogelijke typen vragen (gedwongen keuze, meerdere antwoorden mogelijk, vragen met schaalwaarden, open vragen etc.). Per vraag kun je aangeven of beantwoording verplicht is. De lay-out ziet er standaard verzorgd uit en vaak zijn er nog allerlei mogelijkheden deze aan te passen. De resultaten zijn direct zichtbaar in de vorm van frequentietellingen. De ruwe data kan worden geëxporteerd naar Excel of SPSS voor verdere analyse. Voor beperkte doeleinden (beperkt aantal respondenten, beperkt aantal vragen, beperkte looptijd) zijn sommige pakketten zelfs gratis beschikbaar. Wie beschikt over e-mailadressen van de beoogde respondenten, kan een bericht sturen met een link naar de vragenlijst. Wie daar niet over beschikt, kan de vragenlijst verspreiden door een link te plaatsen op een website, sociaal mediakanaal, discussieforum of nieuwsgroep.

Er kunnen omstandigheden zijn om te kiezen voor schriftelijke of mondelinge afname. Bijvoorbeeld:

- Een onderzoeker heeft de gelegenheid om een vragenlijst aan scholieren of studenten uit te delen tijdens de les. In het lokaal bevinden zich geen of onvoldoende computers om online afname mogelijk te maken.

- Een gedeelte van de doelgroep is mogelijk minder computervaardig en zou dus met een online vragenlijst niet bereikt worden. Dit is met name het geval bij ouderen of laagopgeleiden.
- De onderzoeker verwacht een hogere respons bij schriftelijke of telefonische afname. Om die reden heeft de Koninklijke Bibliotheek er tot voor kort voor gekozen het periodieke gebruikersonderzoek te laten verlopen via een telefonische enquête. Een extra argument was dat mensen die niet wilden meewerken er vaak nog wel toe verleid konden worden om althans een rapportcijfer voor de fysieke dienstverlening en website toe te kennen.

7 Omvang en samenstelling van de steekproef

Tot nu toe hebben we ons vooral gericht op WAT en HOE we het gaan vragen. Deze paragraaf gaat in op WIE we gaan bevragen. Gebruikersonderzoek richt zich meestal niet op de gehele populatie, maar op een steekproef daaruit. Uit de definitie van survey, eerder opgenomen in paragraaf 1, valt op te maken dat een onderzoek pas als survey gekenschetst kan worden als dit verloopt via een steekproef. Onderzoek waarbij de gehele populatie wordt betrokken wordt aangeduid met de term census, in het Nederlands vertaald als volkstelling. Dit duidt erop dat de term census vooral geassocieerd wordt met onderzoek onder alle inwoners van een land of regio.

Enkele decennia geleden was het praktisch vaak niet uitvoerbaar om de gehele populatie bij het onderzoek te betrekken. De Universiteit van Amsterdam bijvoorbeeld telt ruim 30.000 studenten. Deze zouden allemaal aangeschreven moeten worden en de onderzoeker zou alle schriftelijke antwoordformulieren handmatig moeten verwerken. Dan is een steekproef van 400 of 500 studenten veel meer hanteerbaar. De praktische bezwaren van toen gelden nu niet meer. In principe is het zeer wel mogelijk alle studenten te benaderen met een e-mail, inclusief een link naar een online vragenlijst. Bij de analyse maakt het weinig uit of deze nu betrekking heeft op 400 of 30.000 studenten. Toch zien we daarvan af, om enquêtemoetheid zoveel mogelijk te voorkomen.

7.1 Uitspraken

Uiteindelijk zijn we niet geïnteresseerd in uitspraken over de steekproef, maar willen we uitspraken doen over de populatie. Dat kan ook, maar met twee slagen om de arm. Stel dat 60% van de studenten uit de steekproef een bijbaantje heeft, of dat de studenten uit de steekproef gemiddeld 28 uur per week aan hun studie besteden. Doorgetrokken naar de populatie kunnen we hierover nu de volgende uitspraken doen:

- We zijn er voor 95% (of 99%) zeker van dat 60% van alle studenten, plus of min een bepaalde marge, een bijbaantje heeft. Die marge kunnen we uitrekenen. De onderzoeker zou nu bijvoorbeeld de volgende conclusie kunnen trekken: met een zekerheid van 95% geldt dat 55,8% tot 64,2% van de studenten een bijbaantje heeft.
- We zijn er voor 95% (of 99%) zeker van dat studenten gemiddeld 28 uur per week aan hun studie besteden, plus of min een bepaalde marge. Die marge kunnen we weer uitrekenen. Dan kunnen we een uitspraak doen als: met een zekerheid van 95% geldt dat studenten gemiddeld tussen 24,1 en 31,9 uur per week aan hun studie besteden.

Zoals gezegd houden we twee slagen om de arm. De uitspraak bevat een *betrouwbaarheidsniveau* en een *betrouwbaarheidsinterval*. Een betrouwbaarheidsniveau van 95% of 99% is conventie. Een waarde van 80% is te laag; er is dan 20% kans dat we ernaast zitten. Soms worden waarden als 90% en 99,9% gehanteerd. Dit laatste komt vooral voor bij medisch onderzoek: als het gaat om de gezondheid van mensen is een zo hoog mogelijke mate van betrouwbaarheid vereist. Bij de sociale wetenschappen en ook op het gebied van informatiewetenschap is een betrouwbaarheidsniveau van 95% gebruikelijk.

Het betrouwbaarheidsinterval willen we zo klein mogelijk houden. We hebben weinig aan een uitspraak als 'we zijn er voor 95% zeker van dat 37% tot 83% van de studenten een bijbaantje heeft', of dat de gemiddelde studietijd ergens tussen 8 en 48 uur ligt. De marge hangt onder meer, maar niet uitsluitend, af van de omvang van de steekproef. Bij een steekproef van 40 studenten moeten we een enorme slag om de arm maken. Bij een steekproef van 400 studenten verkleinen we de onzekerheidsmarge sterk.

Op ons vakgebied volstaat vaak een steekproef van circa 400 personen. Beter gezegd: van 400 daadwerkelijk responderende personen. Het is

een wijdverbreid misverstand dat de omvang van de steekproef in verhouding moet staan tot de omvang van de populatie. Of de populatie nu bestaat uit 3.000, 30.000 of 300.000 personen maakt weinig uit voor de benodigde omvang van de steekproef. In alle gevallen levert een steekproef van 400 personen nagenoeg dezelfde marge van onzekerheid op. Pas bij hele kleine populaties kunnen we met een minder grote steekproef volstaan. Binnenkort verschijnt in het *Handboek Informatiewetenschap* een bijdrage over inductieve statistiek, waarin nader ingegaan wordt op deze en aanverwante vraagstukken.

7.2 Typen steekproeven

Hoe gaan studenten van de UvA om met Google Scholar? Om dit te onderzoeken, besluiten we een steekproef te nemen van 400 studenten. Niet alleen de omvang, maar ook de samenstelling van de steekproef vereist onze aandacht. We zouden twee wegen kunnen bewandelen. Uit de administratie van de universiteit selecteren we, geheel willekeurig, 400 studenten; of we wachten studenten op bij het verlaten van collegezalen en gaan net zolang door tot we er 400 hebben. In het eerste geval hebben we te maken met een aselechte (random) steekproef, in het tweede geval met een selecte. Bij een aselechte steekproef heeft ieder lid van de populatie, in dit geval elke UvA student, evenveel kans om in de steekproef terecht te komen. Bij een selecte steekproef is dat niet het geval. Daaraan is een groot nadeel verbonden: mogelijk is de steekproef niet representatief voor de gehele populatie. In ons voorbeeld zouden we waarschijnlijk veel studenten van eenzelfde studierichting trekken; mogelijk zijn studenten die daadwerkelijk colleges bijwonen al niet representatief voor alle studenten.

Figuur 2 geeft een overzicht van verschillende soorten steekproeven.

| Aselechte steekproeven | Selecte steekproeven |
|---------------------------------------|---|
| A1. Enkelvoudige aselechte steekproef | B1. Gelegenheidssteekproef (gemakssteekproef) |
| A2. Systematische steekproef | B2. Quotasteekproef |
| A3. Gestratificeerde steekproef | B3. Doelgerichte steekproef |
| A4. Getrapte steekproef | B4. Sneeuwbalsteekproef |
| A5. Clustersteekproef | B5. Vrijwillige deelname/zelfselectie |

Figuur 2. Typologie van steekproeven.

- A1. Bij een *enkelvoudige aselechte steekproef* is het alsof we alle leden van een populatie in een hoge hoed gooien en daar volstrekt willekeurig een aantal personen uit kiezen. Als je een lijst hebt van alle personen, kun je daaruit zowel met Excel als SPSS een aselechte steekproef trekken.
- A2. Een *systematische steekproef* gaat ook uit van een lijst van alle leden van de populatie. Als we 400 personen nodig hebben uit een bestand van 30.000, dan nemen we er een op de 75. We beginnen bijvoorbeeld met persoon 17 (aselect begin), en kiezen vervolgens persoon 92, 167, 242, 317 enzovoort. Voorwaarde is dat er geen systematiek in de lijst is, de personen vooraan niet afwijken van die elders in de lijst. Een variant is elke tiende bezoeker van bijvoorbeeld een studiezaal voor het onderzoek te benaderen. In de tijd van de kaartcatalogus trokken we om de zoveel centimeter een kaartje.
- A3. Je kiest voor een *gestratificeerde steekproef* als je verzekerd wilt zijn van een goede afspiegeling van subgroepen. Stel dat een populatie bestaat uit de volgende leeftijdscategorieën:

| Leeftijdscategorie | Aantal personen in populatie | Aantal personen in proportionele gestratificeerde steekproef | Aantal personen in disproportionele gestratificeerde steekproef |
|--------------------|------------------------------|--|---|
| 30 of jonger | 1.000 (4%) | 16 (4%) | 64 |
| 31 – 40 jaar | 5.000 (20%) | 80 (20%) | 80 |
| 41 – 50 jaar | 10.000 (40%) | 160 (40%) | 160 |
| 51 – 60 jaar | 8.000 (32%) | 128 (32%) | 128 |
| 61 jaar of ouder | 1.000 (4%) | 16 (4%) | 64 |
| Totaal | 25.000 (100%) | 400 (100%) | 496 |

Als leeftijd een belangrijke variabele in het onderzoek is, kunnen we besluiten om in de steekproef eenzelfde verhouding tussen leeftijdscategorieën aan te houden. Het resultaat is weergegeven in de derde kolom. We zien dat er twee groepen zijn met slechts 16 personen in de steekproef. In de vierde kolom hebben we die bewust oververtegenwoordigd. We spreken dan niet langer van een proportionele, maar van een

disproportionele gestratificeerde steekproef. Voor het gemak hebben we de steekproefomvang van de andere categorieën niet verminderd, zodat de totale omvang stijgt van 400 naar 496. Het voordeel is dat uitspraken over die subgroepen op een meer aanvaardbare steekproefomvang berusten. Bij het berekenen van de resultaten voor de volledige steekproef moeten we wel een correctie toepassen: de resultaten van de jongste en oudste leeftijdsgroep krijgen elk een gewicht van $1/4$. Onderstaand voorbeeld toont hoe een dergelijke correctie in zijn werk gaat. Daarbij is tevens rekening gehouden met een response die beduidend lager ligt dan 100%. Na correctie bedraagt het percentage 'ja' niet 57,8%, maar 51,6%.

| | Ongecorrigeerde resultaten | | | Gecorrigeerde resultaten | | | |
|---------|----------------------------|------------|------------|--------------------------|------------|------------|-------------|
| | Ja | Nee | Totaal | | Ja | Nee | Totaal |
| < 31 | 32 | 8 | 40 | < 31 | $32/4 = 8$ | $8/4 = 2$ | $40/4 = 10$ |
| 31 - 40 | 24 | 19 | 43 | 31 - 40 | 24 | 19 | 43 |
| 41 - 50 | 30 | 41 | 71 | 41 - 50 | 30 | 41 | 71 |
| 51 - 60 | 28 | 26 | 54 | 51 - 60 | 28 | 26 | 54 |
| > 60 | 20 | 4 | 24 | > 60 | $20/4 = 5$ | $4/4 = 1$ | $24/4 = 6$ |
| Totaal | 134 (57,8%) | 98 (42,2%) | 232 (100%) | | 95 (51,6%) | 89 (48,4%) | 184 (100%) |

Stratificatie vereist wel dat we op voorhand op de hoogte zijn van kenmerken zoals leeftijd, geslacht en studierichting. Naarmate steekproeven groter zijn, neemt de kans op een evenredige verdeling automatisch toe en is het minder nodig om tot stratificatie over te gaan.

- A4. Stel dat we een onderzoek willen verrichten, niet onder de studenten van de UvA, maar onder alle studenten in Nederland. We zouden dan over een lijst van alle studenten aan Nederlandse universiteiten moeten beschikken. Het kan praktischer zijn om eerst enkele universiteiten te kiezen, en daarbinnen de studenten. De eerste trap is dan een steekproef van universiteiten, de tweede van studenten binnen de gekozen instellingen. Een bekende vorm van *getrapte steekproeven* is 'area sampling'. We doen een onderzoek dat zich richt op alle inwoners van Nederland en selecteren eerst een aantal provincies, daarbinnen gemeenten, daarbinnen misschien nog wijken en straten, en uiteindelijk de bewoners. Hoe meer trappen, des te groter de kans dat de steekproef aan

representativiteit verliest. De enige reden om te kiezen voor een getrapte steekproef is van praktische aard.

- A5. Een variant op de getrapte steekproef is de *clustersteekproef*. We willen een uitspraak doen over de mediawijsheid van alle kinderen in groep 6 van de basisschool. Ook nu weer trekken we eerst een aselechte steekproef van scholen. Binnen elke school nemen we nu alle kinderen van groep 6, niet een steekproef van kinderen uit groep 6. Deze klassen zijn bestaande eenheden, die in dit kader als cluster aangeduid worden.

Lang niet altijd beschikt de onderzoeker over een lijst van alle leden van de populatie. Zo'n lijst wordt ook wel steekproefkader genoemd. Noodgedwongen moet hij zich dan behelpen met een selecte steekproef. Er is geen enkele garantie dat deze representatief is voor de gehele populatie. Het is daarom ook niet goed mogelijk de resultaten te generaliseren naar de gehele populatie. De resultaten zijn hooguit indicatief.

- B1. De termen *gelegenheid of gemak*, of de Angelsaksische termen *accidental, grab, convenience* of *opportunity* geven goed aan waar het bij deze wijze van steekproeftrekking om gaat. De onderzoeker betreft mensen uit zijn eigen kennissenkring of werkomgeving bij het onderzoek of spreekt de eerste de beste voorbijgangers op straat aan.
- B2. *Quotasteekproeven* zijn een variant daarop. Extra eis is dat bepaalde 'quota' gehaald moeten worden, bijvoorbeeld 200 mannen en 200 vrouwen. Daarmee kun je ervoor zorgen dat de steekproef althans in één opzicht een weerspiegeling is van de populatie. Anders dan bij stratificatie worden mensen binnen de quota op selectieve wijze geworven.
- B3. Een *doelgerichte steekproef* (*purposive, judgement, selective, subjective sampling*) is gericht op een specifieke doelgroep. Helaas beschikken we weer niet over een steekproefkader. Een doelgerichte steekproef kan daarom de vorm aannemen van een *gelegenheidssteekproef*: pik er op straat alleen die mensen uit die om wat voor reden dan ook tot je doelgroep lijken te behoren. Hetzelfde effect bereik je door mensen te benaderen op specifieke locaties, zoals musea of

onderwijsgebouwen. Als je hen aanspreekt in hun hoedanigheid van museumbezoeker of student, heb je te maken met een doelgerichte steekproef. Als je hen aanspreekt als vertegenwoordigers van een veel bredere categorie zoals gebruikers van online nieuwsmedia of bezoekers van supermarkten, is eerder sprake van een gelegenheidssteekproef.

- B4. Ook bij een *sneeuwbalsteekproef* hebben we te maken met een specifieke doelgroep, waar we helaas geen lijst van hebben. Hopelijk kennen we wel een of enkele vertegenwoordigers, vraag hen of zij nog soortgenoten kennen. Via via komen we zo verder. Eigenlijk hebben we hier te maken met een doelgerichte steekproef, met dat verschil dat de leden op een bijzondere wijze geworven worden. Deze methode is toegepast door Saskia Leefsma in haar onderzoek naar de documentatie van werken van overleden kunstenaars.
- B5. De meest voorkomende vorm van selecte steekproeven in de wereld van erfgoedinstellingen is *zelfselectie*, ofwel deelname op vrijwillige basis. De onderzoeker plaatst een link naar een vragenlijst op een website of forum. Er bestaat helaas geen lijst van mensen die een website als Delpher bezoeken; om de bezoekers te benaderen zijn er nauwelijks andere mogelijkheden dan een vragenlijst via de website aan te bieden. Het nadeel daarvan is dat er geen enkel zicht is op wie de vragenlijst beantwoordt en daarmee op de representativiteit. Mogelijk zullen juist de enthousiaste gebruikers, mensen met een uitgesproken mening of mensen met meer vrije tijd de gelegenheid aangrijpen hun stem te laten horen. Daarom moeten we weer heel voorzichtig zijn met het generaliseren van de resultaten, zeker als de response laag is.

Een onderzoek van Harley en Henke (2007) was in dit opzicht verhelderend. Zij plaatsten op twee websites een online vragenlijst. De response was 161, respectievelijk 45 personen. Om een indruk te krijgen van de hoogte en de representativiteit van de response analyseerde men de webstatistieken. Aldus kon men vaststellen dat tijdens de enquêteperiode de sites zijn geraadpleegd door 97.284 respectievelijk 38.962 verschillende personen. Het aantal respondenten ligt in beide onderzoeken dus lager dan 0,2%. Dit alleen al geeft aan dat de resultaten allerminst

representatief zijn. In tabel 2 is bovendien duidelijk te zien dat de respondenten bovengemiddeld intensieve gebruikers zijn, zowel gemeten naar het aantal sessies als naar de tijdsduur van deze sessies. Zij zijn daarmee althans in dit opzicht niet representatief voor de gehele bezoekerspopulatie. Alle resultaten van het onderzoek moeten dus met de nodige korrels zout genomen worden.

| | Website 1 | | Website 2 | |
|--------------------------------------|---------------------------|--------------------------------|--------------------------|--------------------------------|
| | Respondenten (n = 161) | Alle bezoekers (n = 97.284) | Respondenten (n = 45) | Alle bezoekers (n = 38.962) |
| Gemiddeld aantal sessies | 2,2 | 1,5 | 5,0 | 1,4 |
| Gemiddeld aantal seconden per sessie | 24 | 9 | 63 | 11 |

Tabel 2. Gebruik van twee websites door respondenten en alle bezoekers.

Een onderzoek van Skov (2013) toont aan dat de plek waarop je de vragenlijst zet sterk bepalend kan zijn voor het type mensen dat je daarmee aantrekt. Zij richtte zich op online bezoekers van het National Museum of Military History in Kopenhagen. Omdat zij vooral geïnteresseerd was in de bezoekers van de content, plaatste zij de oproep voor deelname aan de enquête niet op de homepage van de website, maar op een pagina die toegang biedt tot de (online) collectie. Daarmee verkleinde zij de kans dat bezoekers die alleen praktische informatie zoeken, zoals openingstijden of entreegelden, bij het onderzoek betrokken worden. Inderdaad bleek dat uit de resultaten. Van de 153 respondenten gaven er 132 een inhoudelijke reden als doel van hun bezoek op. Zeer waarschijnlijk zou plaatsing op de homepage een lager aantal hebben opgeleverd.

7.3 Panels

Bij aselecte steekproeven beschikt de onderzoeker over een steekproefkader, een lijst van alle leden van de populatie. Bij selecte steekproeven is dat niet het geval en moet de onderzoeker zich in bochten wringen om een steekproef te kunnen samenstellen: mensen plukken uit de eigen werkomgeving of kennissenkring, mensen aanspreken op bepaalde locaties, mensen benaderen via een forum of website. Tegenwoordig zijn er ook veel onderzoeksbureaus die panels hebben samengesteld die

voor tal van onderzoeken kunnen worden ingezet. Van de panelleden is vaak een groot aantal achtergrondgegevens bekend, zoals koopgedrag en omgang met media. Dat maakt het mogelijk op maat gesneden groepen samen te stellen.

TNS NIPO is zo'n onderzoeksbureau. Het beschikt over een panel van ruim 200.000 respondenten. Van deze respondenten is standaard een groot aantal gegevens beschikbaar, zoals socio-demografische kenmerken, waardenoriëntatie, informatie over mediaconsumptie, merkgebruik en surfgedrag op internet. De panelleden worden maximaal drie keer per maand benaderd en ontvangen altijd een beloning voor hun medewerking. Dit resulteert in een zeer hoge respons van gemiddeld circa 70%.³

Speciaal voor openbare bibliotheken heeft ProBiblio een online klantenpanel opgezet.⁴ Biebpanel (www.biebpanel.nl) telt ruim twintigduizend deelnemers uit heel Nederland. Zij nemen hooguit vier maal per jaar deel aan een onderzoek. Deelnemende bibliotheken kunnen zelf mede de onderwerpen bepalen. Zij betalen ongeveer 4.000 euro per jaar.

Steekproeven gebaseerd op een panel behoren tot de selecte steekproeven. De leden zijn immers niet rechtstreeks gekozen uit een bepaalde populatie. Het feit dat mensen zich zelf kunnen aanmelden voor een panel kan al een vertekening teweegbrengen. Niet uitgesloten is dat zich daaronder enkele professionele invullers bevinden, die dit als een bijverdienste zien. Naarmate meer achtergrondgegevens van de panelleden bekend zijn, zijn we beter in staat om een voor een bepaald onderzoek op maat gesneden selectie uit het panelbestand te maken. Dit geldt zeker voor specifieke panels, zoals Biebpanel.

De term panelonderzoek bestaat al langer en kan verwarrend zijn. Oorspronkelijk was de term verbonden met longitudinaal onderzoek, waarbij op verschillende tijdstippen min of meer dezelfde vragen worden voorgelegd aan min of meer dezelfde groep mensen. Het doel is om trends vast te stellen. Een dergelijke vorm van panelonderzoek bestond al lang voor het online tijdperk.

8 Analyse van de resultaten

Dit hoofdstuk gaat in op de problematiek van lage en mogelijk selectieve response (8.1), enkele basisbegrippen uit de beschrijvende statistiek (8.2) en de vraag of met gebruikersonderzoek causale uitspraken mogelijk zijn (8.3).

8.1 Response

In oude handboeken werd wel gesteld dat de response minstens zo'n 75% moet zijn om betrouwbare uitspraken te kunnen doen. Aan deze voorwaarde voldoen tegenwoordig maar weinig onderzoeken. Het is daarom goed daar op voorhand rekening mee te houden. Als je 400 responses wilt behalen, moet de steekproefomvang niet 400, maar eerder 1.000 of nog meer bedragen. Alleen bij objecten kun je er zeker van zijn dat je een response behaalt van 100%: boeken, titelrecords of gedigitaliseerde pagina's zijn niet in staat hun medewerking te ontzeggen.

Ook al behalen we met een vergrote steekproef misschien het benodigde aantal responses in absolute zin, het responsepercentage blijft dikwijls laag. Als 400 op de 1.600 benaderde mensen reageert, dan betekent dat evengoed een non-response van 75%. Wellicht verschillen de niet-responderenden van de wel-responderenden. Vooral degenen die geïnteresseerd zijn in het onderwerp en degenen die denken dat ze invloed kunnen uitoefenen met hun deelname zullen geneigd zijn te reageren. De resultaten van het onderzoek, gebaseerd op de antwoorden van de feitelijke respondenten, leveren daarmee een beeld op dat mogelijk niet representatief is voor de gehele populatie. We hebben wel ons best gedaan een aselechte steekproef samen te stellen, het is echter maar zeer de vraag in hoeverre dat aselechte karakter ook opgaat voor de feitelijk responderende personen.

Strikt genomen zouden we alleen maar uiterst voorzichtige uitspraken kunnen doen. Onderstaande tabel geeft een voorbeeld van enkele resultaten uit een onderzoek. Er zijn 1000 vragenlijsten verstuurd, de response bedraagt 500 (50%). Een van de vragen is of men gebruik maakt van een bepaalde voorziening. Daarop antwoorden 350 respondenten met ja, 150 met nee. We concluderen dat $350/500 = 70%$ van de respondenten de voorziening gebruikt. Het is theoretisch echter denkbaar dat alle 500 niet-responderenden geen gebruik maken van de voorziening, of alle 500 juist wel. Wat we met zekerheid kunnen zeggen is

dat 35% tot 85% van de steekproef gebruiker is. Geen enkele onderzoeker zal zo redeneren, maar het valt niet uit te sluiten.

| | |
|-----------------------|------|
| Respons: 500, waarvan | |
| – wel gebruiker: | 350 |
| – niet gebruiker: | 150 |
| Non respons: | 500 |
| Totaal: | 1000 |

Burkell (2003) geeft enkele suggesties hoe men kan omgaan met het probleem van non-response.

1. Kijk of respondenten en niet-respondenten overeenkomen, althans in die opzichten waarvan je de gegevens hebt. Mogelijk kan dat aanleiding zijn om een gewogen score te berekenen. Onderstaande tabel geeft een voorbeeld. We weten hoeveel personen we benaderd hebben op alfa, gamma en bèta-terrein. Het valt op dat slechts weinig personen op alfagebied de vragenlijst beantwoord hebben en tevens dat degenen die dat wel gedaan hebben weinig gebruik maken van de voorziening. Uit de berekening van de gewogen score in de laatste kolom blijkt dat we misschien beter kunnen spreken van 57,5% gebruik dan 70% gebruik.

| | Steekproef | Response | Gebruik | Gewogen score |
|--------|------------|-----------|-----------|-----------------------|
| Alfa | 400 | 100 (25%) | 20 (20%) | $400/100 * 20 = 80$ |
| Gamma | 300 | 200 (67%) | 170 (85%) | $300/200 * 170 = 255$ |
| Bèta | 300 | 200 (67%) | 160 (80%) | $300/200 * 160 = 240$ |
| Totaal | 1000 | 500 (50%) | 350 (70%) | 575 (57,5%) |

2. Kijk of latere respondenten overeenkomen met eerdere respondenten, althans in eigenschappen waarover je gegevens hebt. De veronderstelling is dat latere respondenten meer dan snelle respondenten gelijkenis vertonen met niet-respondenten. Hopelijk blijken er geen verschillen te zijn tussen vroege en late respondenten.
3. Benader een groep niet-respondenten en probeer hen toch een of twee kernvragen te laten beantwoorden. Al eerder is vermeld dat de KB voorheen telefonische enquêtes hield onder haar leden. Aan degenen die niet wilden meewerken is één kernvraag voorgelegd: een rapportcijfer over de dienstverlening. Zo kon men nagaan of de

(bijna) niet-respondenten afwijken van de wel-respondenten. Gelukkig bleek dat niet het geval te zijn, de gemiddelde rapportcijfers van beide groepen waren even hoog. Grote verschillen zouden wel tot voorzichtigheid manen: mogelijk zijn de resultaten van het onderzoek niet betrouwbaar.

8.2 Basisbegrippen uit de beschrijvende statistiek

Er zijn twee soorten statistiek: beschrijvende en inductieve statistiek. Het resultaat van ons onderzoek is een spreadsheet in Excel of SPSS. De verwerking van deze ruwe gegevens tot overzichtelijke resultaten is het terrein van de beschrijvende statistiek. Hier gaan we beknopt in op tabellen, grafieken, centrummaten, spreidingsmaten en correlatiematen. Daaraan voorafgaand eerst nog een overzicht van meetniveaus. Het is geenszins de bedoeling een uitputtend overzicht te geven. Er is een keur aan statistiekboeken op de markt. Bij inductieve statistiek draait het altijd om de vraag: wat betekenen de gegevens van mijn steekproef voor de populatie? Het begrip significantie speelt daarin een centrale rol. Een afzonderlijke bijdrage in dit handboek is gewijd aan inductieve statistiek.

1. Meetniveaus

Het resultaat van ons onderzoek is een spreadsheet in Excel of SPSS. In de kolommen de variabelen, in de rijen de cases of respondenten, in de cellen de scores in de vorm van cijfers. Met deze cijfers gaan we nu aan de slag. Wat we daarmee kunnen doen, hangt af van het meetniveau. We maken een onderscheid tussen vier categorieën:

1. Nominaal. Het getal is slechts een code, label, naam (nomen): 1 = man, 2 = vrouw. Of 1 = alfa, 2 = gamma, 3 = bèta. Zij hebben geen rekenkundige betekenis en zouden net zo goed anders kunnen zijn: 1 = vrouw, 2 = man. Of 1 = bèta, 2 = gamma, 3 = alfa. We kunnen niet zeggen dat label 3 hoger of beter is dan label 2.
2. Ordinaal. De getallen geven een orde, rangorde weer. Daarvan is sprake bij schaalwaarden zoals 1 = heel ontevreden, 2 = ontevreden, 3 = neutraal, 4 = tevreden, 5 = heel tevreden. We kunnen nu wel zeggen dat 4 hoger is dan 3. Maar niet dat 4 twee keer zo goed is als 2. De onderlinge afstanden tussen de schaalwaarden zijn onbekend en kunnen sterk verschillen.
4. Ratio. We hebben nu te maken met echte getallen, waarop we alle rekenkundige bewerkingen kunnen uitvoeren. Voorbeelden zijn leeftijd, collectieomvang, bezoekersaantallen, downloadfrequentie.

We kunnen zeggen dat 40 jaar twee keer zo oud is als 20 jaar, en dat 10.000 bezoekers vijf keer zoveel is als 2.000 bezoekers.

3. Interval. Deze hebben we nog even overgeslagen omdat we hier maar zelden mee te maken hebben. De bekendste intervalvariabelen zijn IQ en temperatuur. Er zijn vaste intervallen: het verschil tussen 15 en 10 graden is even groot als het verschil tussen 20 en 15 graden. Maar er is geen vast nulpunt. We kunnen dus niet zeggen dat 20 graden twee keer zo warm is als 10 graden. Of dat een IQ van 140 twee keer zo hoog is als een IQ van 70.

2. Tabellen

Tabel 3 is een voorbeeld van een enkelvoudige frequentietabel, tabel 4 van een kruistabel. De gegevens zijn afkomstig van een onderzoek naar literatuurverwijzingen in 236 papers van studenten (Mill, 2008).

| | (n=236) | % |
|----------|---------|-------|
| Journals | 1.153 | 47,6% |
| Books | 726 | 29,9% |
| Websites | 406 | 16,7% |
| Other | 140 | 5,7% |
| Total | 2.425 | 100% |

Tabel 3. Enkelvoudige tabel: soort literatuurverwijzingen in papers.

| | Humanities (n=63) | Sciences (n=64) | Social sciences (n=109) | All divisions (n=236) |
|----------|----------------------|--------------------|-------------------------------|--------------------------|
| Journals | 114 (24,5%) | 421 (66,2%) | 618 (46,7%) | 1.153 (47,6%) |
| Books | 283 (60,7%) | 110 (17,3%) | 333 (25,2%) | 726 (29,9%) |
| Websites | 49 (10,5%) | 95 (14,9%) | 262 (19,8%) | 406 (16,7%) |
| Other | 20 (4,3%) | 10 (1,6%) | 110 (8,3%) | 140 (5,7%) |
| Totals | 466 (100%) | 636 (100%) | 1.323 (100%) | 2.425 (100%) |

Tabel 4. Kruistabel: soort literatuurverwijzingen in papers, per discipline.

Het voorbeeld laat zien dat er een duidelijk verband is tussen discipline en de keuze van documentsoorten. Zonder uitsplitsing van de resultaten naar vakgebied zouden deze verborgen gebleven zijn. We maken deze

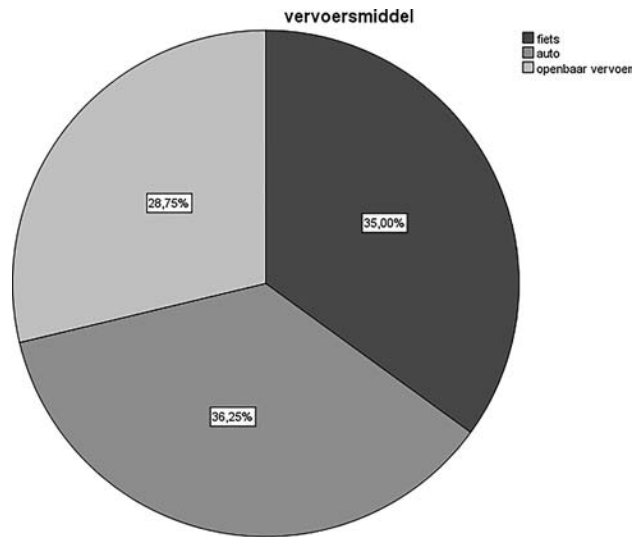
uitsplitsing omdat we een verband vermoeden. In feite hanteren we een impliciet of misschien wel expliciet conceptueel model.

De conventie is om oorzaakvariabelen in de kolommen te plaatsen, gevolgvariabelen in de rijen. De studierichting is bepalend voor welk soort bronnen men gebruikt, niet omgekeerd. Om praktische redenen kan men daar soms van afwijken. Als we tien disciplines zouden onderscheiden en nog steeds vier documentsoorten, zouden we een tabel krijgen met tien kolommen en vier rijen. Ter wille van de overzichtelijkheid draaien we dat liever om.

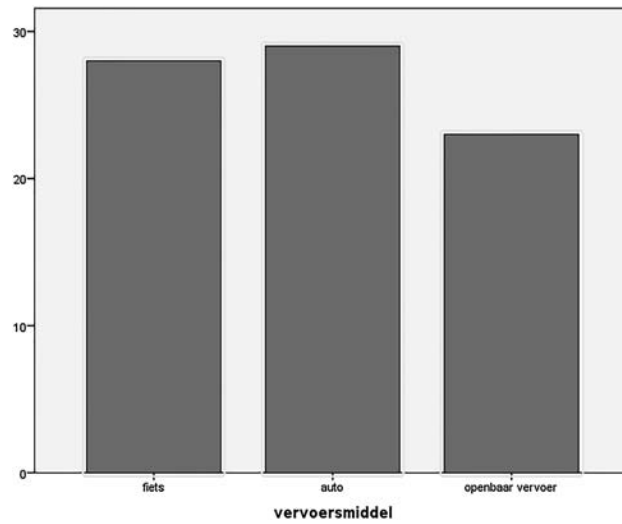
3. *Grafieken*

Je kunt de frequenties in een tabel, maar ook in een grafiek weergeven. Figuur 3a toont de verdeling van vervoersmiddelen van de respondenten in de vorm van een cirkel- of taartdiagram. Een alternatief is de staafdiagram of bar chart (figuur 3b). Het is een kwestie van smaak welke van beide men verkiest. In de figuren 3c en 3d is nog een uitsplitsing gemaakt van vervoersmiddelen naar geslacht. Dit kan op twee manieren: clustering of stapeling.

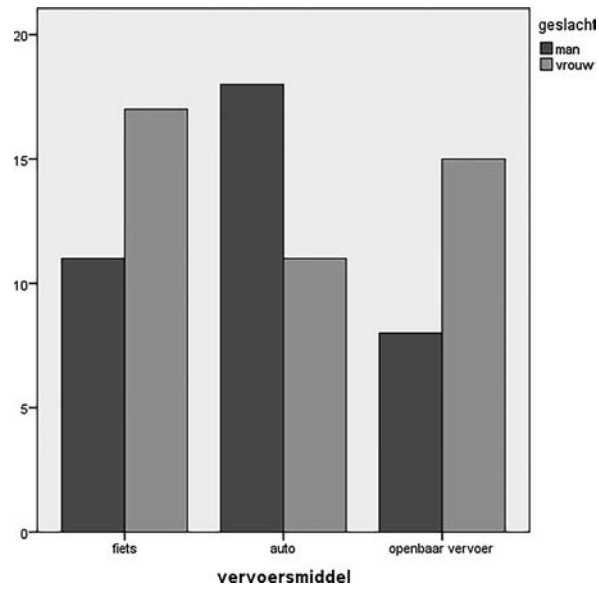
Bij staafdiagrammen is er een lege ruimte tussen de staafjes. Bij een histogram (figuur 3e) liggen deze staafjes tegen elkaar aan. Deze passen we toe bij ratiovariabelen. Er is geen logische, aaneengesloten volgorde tussen vervoersmiddelen, wel tussen reisafstanden.



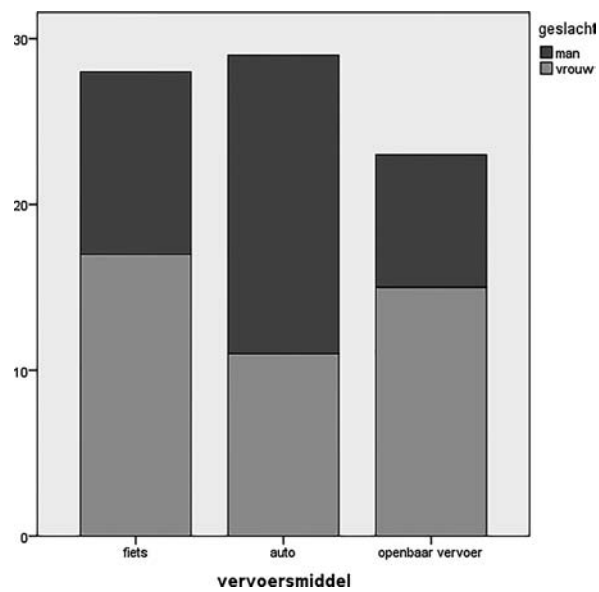
a. *Cirkel- of taartdiagram (pie chart): verdeling naar vervoersmiddel*



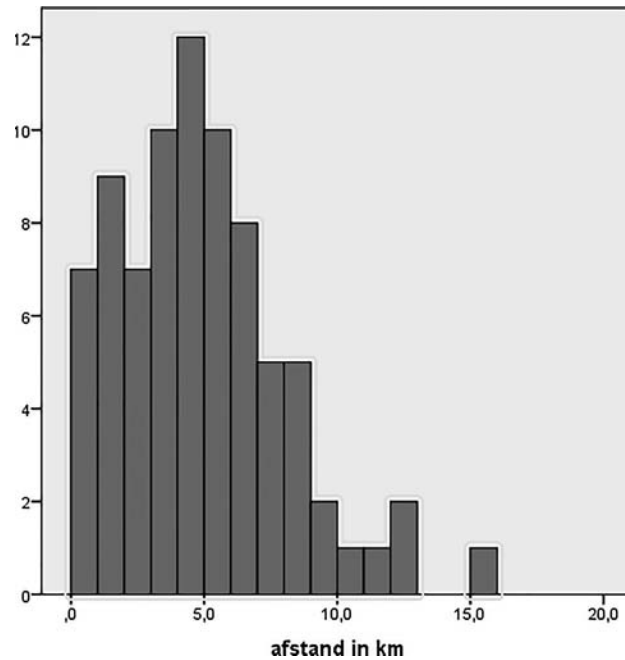
b. *Staafdiagram (simple bar chart): verdeling naar vervoersmiddel*



c. Clustered bar chart: vervoersmiddel, onderverdeeld naar man/vrouw



d. Stacked bar chart: vervoersmiddel, onderverdeeld naar man/vrouw



e. *Histogram: verdeling naar reisafstand*

Figuur 3. Typen grafieken.

4. *Centrummaten*

We verrichten een meting bij vijftien personen. De scores, op een rijtje gezet, zijn:

5, 7, 7, 8, 10, 10, 10, **12**, 12, 14, 16, 17, 18, 20, 44

Het rekenkundig gemiddelde bedraagt $210/15 = 14,0$. In dit geval geeft dit niet een goed beeld van de gemiddelde respondent: tien van de vijftien personen hebben een score onder het gemiddelde. Het gemiddelde wordt hier sterk beïnvloed door de extreme waarde 44. Een alternatief is dan om uit te gaan van de mediaan, de middelste score, in dit geval 12 (vet aangegeven). De mediaan is minder gevoelig voor uitschieters, extreme scores. Een derde optie is de modus: de meest voorkomende waarde. In het voorbeeld is de modus 10, deze score komt drie keer voor.

Het is mogelijk dat er meer dan één modus is. Als we dat grafisch afbeelden zien we twee of meer toppen en spreken we van een bimodale of multimodale verdeling.

Het rekenkundig gemiddelde mogen we alleen gebruiken bij ratiovariabelen. Toegepast op nominale variabelen zou je onzinnige uitspraken krijgen zoals ‘het gemiddelde geslacht is 1,37’. Eigenlijk mag je ook geen gemiddelden berekenen van ordinale variabelen, maar in de praktijk wordt dit veel gedaan omdat iedereen zich daar gemakkelijk een voorstelling van kan maken: ‘de gemiddelde schaalwaarde is 3,65’.

5. *Spreidingsmaten*

Het gemiddelde zegt niet alles. De twee onderstaande reeksen hebben beide een gemiddelde van 14,0, maar verschillen sterk in spreiding:

5, 7, 7, 8, 10, 10, 10, 12, 12, 14, 16, 17, 18, 20, 44

12, 12, 13, 13, 13, 14, 14, 14, 14, 14, 15, 15, 15, 16, 16

De meest gehanteerde spreidingsmaat is de standaarddeviatie (s). Deze wordt als volgt berekend:

1. Bepaal het gemiddelde m .
2. Bereken het verschil van elke score x en het gemiddelde m .
3. Kwadrateer elk verschil.
4. Tel de gekwadrateerde scores op (Sum of Squares, ofwel SS).
5. Bereken de variantie (s^2) door SS te delen door het aantal respondenten (n) of het aantal respondenten min 1 ($n - 1$).
6. Bereken de standaarddeviatie (s) door de wortel te trekken uit de variantie.

| <i>Reeks 1: m = 14,0</i> | | <i>Reeks 2: m = 14,0</i> | |
|---|--------------|---|------------|
| $(5 - 14)^2$ | = 81 | $(12 - 14)^2$ | = 4 |
| $(7 - 14)^2$ | = 49 | $(12 - 14)^2$ | = 4 |
| $(7 - 14)^2$ | = 49 | $(13 - 14)^2$ | = 1 |
| $(8 - 14)^2$ | = 36 | $(13 - 14)^2$ | = 1 |
| $(10 - 14)^2$ | = 16 | $(13 - 14)^2$ | = 1 |
| $(10 - 14)^2$ | = 16 | $(14 - 14)^2$ | = 0 |
| $(10 - 14)^2$ | = 16 | $(14 - 14)^2$ | = 0 |
| $(12 - 14)^2$ | = 4 | $(14 - 14)^2$ | = 0 |
| $(12 - 14)^2$ | = 4 | $(14 - 14)^2$ | = 0 |
| $(14 - 14)^2$ | = 0 | $(14 - 14)^2$ | = 0 |
| $(16 - 14)^2$ | = 4 | $(15 - 14)^2$ | = 1 |
| $(17 - 14)^2$ | = 9 | $(15 - 14)^2$ | = 1 |
| $(18 - 14)^2$ | = 16 | $(15 - 14)^2$ | = 1 |
| $(20 - 14)^2$ | = 36 | $(16 - 14)^2$ | = 4 |
| $(44 - 14)^2$ | = <u>900</u> | $(16 - 14)^2$ | = <u>4</u> |
| | 1236 | | 22 |
| <i>a. Standaarddeviatie van volledige populatie</i> | | <i>a. Standaarddeviatie van volledige populatie</i> | |
| Deel Sum of Squares door aantal cases (n) → variantie $s^2 = 1236 / 15 = 82,4$ → standaarddeviatie $s = \sqrt{82,4} = 9,08$ | | Deel Sum of Squares door aantal cases (n) → variantie $s^2 = 22 / 15 = 1,47$ → standaarddeviatie $s = \sqrt{1,47} = 1,21$ | |
| <i>b. Schatting van standaarddeviatie van populatie op basis van steekproef</i> | | <i>b. Schatting van standaarddeviatie van populatie op basis van steekproef</i> | |
| Deel Sum of Squares door aantal cases min 1 (n-1) → variantie $s^2 = 1236 / 14 = 88,29$ → standaarddeviatie $s = \sqrt{88,29} = 9,40$ | | Deel Sum of Squares door aantal cases min 1 (n-1) → variantie $s^2 = 22 / 14 = 1,57$ → standaarddeviatie $s = \sqrt{1,57} = 1,25$ | |

Inderdaad is de standaarddeviatie van de eerste reeks veel groter dan die van de tweede reeks. Standaarddeviaties krijgen pas betekenis door onderlinge vergelijking, de waarden zelf spreken niet echt tot de verbeelding. De hoogte van de standaarddeviatie bevindt zich altijd ergens de hoogste en laagste afwijking. In de eerste reeks dus tussen 0 en 30, in de tweede tussen 0 en 2.

Een complicatie is dat de standaarddeviatie op twee manieren berekend kan worden, afhankelijk van het doel. Als we de volledige populatie onderzocht hebben, delen we de Sum of Squares door n. Als we op basis van onze steekproefgegevens een uitspraak willen doen over de standaarddeviatie van de populatie, delen we de Sum of Squares door $n - 1$. SPSS kiest automatisch voor de tweede optie. In de praktijk maakt het echter weinig uit of we een getal nu delen door 400 of 399. Maar waarom zouden we eigenlijk een schatting willen maken van de standaarddeviatie van de populatie? Omdat dat aan de basis ligt van tal van berekeningen in de inductieve statistiek. Zie weer de afzonderlijke bijdrage over dit onderwerp.

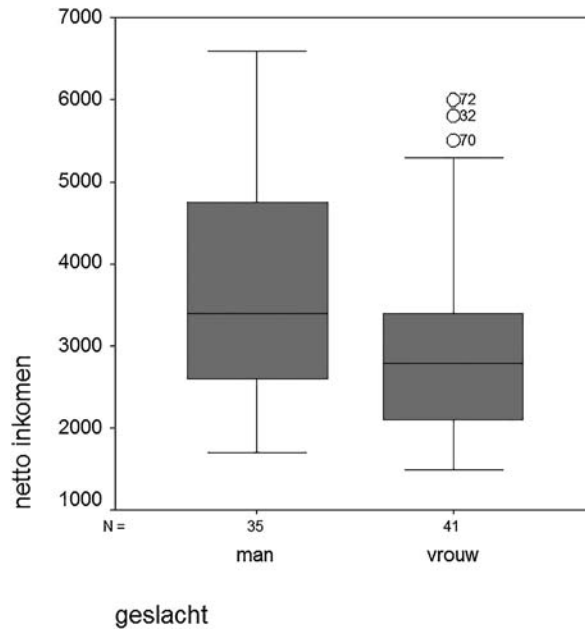
De spreiding wordt grafisch goed in beeld gebracht door een boxplot (doosdiagram), voluit box and whiskerplot. Ook nu weer is dit vooral zinvol bij onderlinge vergelijking. Figuur 4 vergelijkt de spreiding van het inkomen van mannen en vrouwen. De gegevens zijn fictief. Een boxplot wordt als volgt samengesteld:

- De horizontale lijn in de box is de mediaan. Die ligt bij mannen een stuk hoger dan bij vrouwen.
- De box bevat alle respondenten in het tweede en derde kwartiel. We grijpen even terug op de eerdere reeks. De mediaan is de waarde in het midden, en daarmee de begrenzing van het tweede kwartiel. Tevens is nu de begrenzing van het eerste (Q_1) en derde kwartiel (Q_3) afgebeeld. Deze bedragen 8 en 17. In een boxplot zouden de waarden 8 en 17 de onderkant en de bovenkant van de box zijn. Daarin bevinden zich de 50% meest modale gevallen: de gevallen in het tweede en derde kwartiel.

5, 7, 7, $\boxed{8}$, 10, 10, 10, $\boxed{12}$, 12, 14, 16, $\boxed{17}$, 18, 20, 44
 $\boxed{Q_1}$ $\boxed{Q_2}$ $\boxed{Q_3}$

In één oogopslag zien we hoe het inkomen van de 50% meest modale mannen en vrouwen zich verhoudt: de bovengrens, de mediaan, de ondergrens.

- De lijnen naar boven en beneden zijn net zo lang doorgetrokken tot de hoogste en laagste score. Althans tot op zekere hoogte: maximaal 1,5 keer de interkwartielafstand. In de voorbeeldreeks is de interkwartiel afstand $17 - 8 = 9$. De bovengrens zou dan liggen bij maximaal $17 + (1,5 \times 9) = 30,5$. De waarde 44 valt hierbuiten en geldt als uitschieter. De ondergrens zou tot nul kunnen doorlopen. In de figuur zien we dat drie vrouwen (cases 70, 32 en 72) een inkomen hebben dat ver boven dat van de rest uitsteekt.



Figuur 4. Boxplot van inkomens van mannen en vrouwen.

6. Correlatiematen

Figuur 5 is een screenshot met enkele fictieve gegevens. Op het oog zien we al een samenhang, een verband, een correlatie tussen uitleenfrequentie en downloadfrequentie: personen die veel lenen zijn over het algemeen ook de personen die veel downloaden. We hebben hier te maken met twee variabelen op rationiveau. In dat geval kunnen we als maat voor de correlatie 'de Pearson' gebruiken, voluit: Pearson's productmoment correlatiecoëfficiënt. Deze kunnen we gemakkelijk uitrekenen met SPSS. De score is een waarde tussen -1,00 en +1,00. Naarmate de score meer de waarde 1 bereikt, in positieve of negatieve zin, is de relatie sterker. Een positieve waarde betekent dat hogere scores op variabele 1 samengaan met hogere scores op variabele 2. Een negatieve waarde betekent dat hogere scores op variabele 1 samengaan met lagere scores op variabele 2; dit is evengoed een relatie. Een waarde in de buurt van 0,00 betekent dat er geen relatie is: de scores op variabele 1 zeggen

helemaal niets over de scores op variabele 2. Voor het voorbeeld geldt: $r = .460$ of, in Nederlandse notatie: $r = 0,460$. Voor de Pearson gebruiken we de letter r als notatie.

| | uitleningen | downloads | curusbibl | studieuren |
|----|-------------|-----------|-----------|------------|
| 1 | 2 | 30 | ja | 30 |
| 2 | 4 | 42 | ja | 30 |
| 3 | 0 | 10 | nee | 16 |
| 4 | 0 | 0 | nee | 17 |
| 5 | 3 | 6 | nee | 20 |
| 6 | 7 | 2 | ja | 22 |
| 7 | 3 | 5 | nee | 28 |
| 8 | 0 | 2 | nee | 10 |
| 9 | 0 | 2 | nee | 12 |
| 10 | 9 | 26 | ja | 32 |
| 11 | 8 | 27 | ja | 28 |

Figuur 5. Gedeelte van datablad in SPSS.

Als een of beide variabelen op ordinaal niveau is, moeten we een andere maat gebruiken. We hebben de keuze tussen Spearman's rho en Kendall's tau. Ook deze kunnen we weer gemakkelijk uitrekenen met SPSS. Om toch enig idee van de achtergrond te hebben, volgt hier een voorbeeld van de berekening van Spearman's rho. We willen nagaan of er een verband is tussen de prijs en de kwaliteit van tablets. De prijs is een variabele op rationiveau, maar de kwaliteit is gemeten door aan deskundigen rapportcijfers te vragen. Dit is een variabele op ordinaal niveau. We kunnen dus niet de Pearson gebruiken. De gegevens zijn fictief.

In kolom 1 van tabel 5 zien we dat we tien tablets onderzocht hebben. Kolom 2 vermeldt de prijs. In kolom 3 hebben we die op volgorde gezet: de goedkoopste op 1, de duurste op 10. Kolom 4 geeft de rapportcijfers van de deskundigen, in kolom 5 zijn deze weer in rangnummers weergegeven. Kolom 6 geeft van elke tablet het verschil (D van difference) in rangnummer van prijs en kwaliteit. Voor de eerste tablet

is dat $5 - 3 = 2$, voor de tweede $8 - 10 = -2$. De laatste kolom kwadrateert deze verschillen. De som van de gekwadraterde verschillen bedraagt 120. Het aantal tablets (n) = 10.

| 1. Tablet | 2. Prijs | 3. Rang prijs | 4. Oordeel | 5. Rang oordeel | 6. D | 7. D ² |
|-----------|----------|---------------|------------|-----------------|------|--------------------|
| 1 | 200 | 5 | 6,7 | 3 | 2 | 4 |
| 2 | 300 | 8 | 9,0 | 10 | - 2 | 4 |
| 3 | 100 | 1 | 6,9 | 4 | - 3 | 9 |
| 4 | 130 | 2 | 7,9 | 7 | - 5 | 25 |
| 5 | 260 | 7 | 7,3 | 5 | 2 | 4 |
| 6 | 800 | 10 | 8,0 | 8 | 2 | 4 |
| 7 | 250 | 6 | 6,5 | 2 | 4 | 16 |
| 8 | 150 | 3 | 8,7 | 9 | - 6 | 36 |
| 9 | 180 | 4 | 5,8 | 1 | 3 | 9 |
| 10 | 450 | 9 | 7,7 | 6 | 3 | 9 |
| | | | | | | $\Sigma D^2 = 120$ |

Tabel 5. Voorbeeld berekening Spearman's rho.

We vullen nu de formule in en zien een zwakke positieve relatie tussen prijs en kwaliteit. De prijs zegt dus wel iets over de kwaliteit, maar niet veel.

$$\rho = 1 - \frac{6 \Sigma D^2}{n(n^2-1)} = 1 - \frac{6 \times 120}{10(100-1)} = 1 - 0,73 = 0,27$$

NB. We hebben ervoor gekozen om de goedkoopste tablet rangnummer 1 en de duurste rangnummer 10 te geven. We hadden dit ook andersom kunnen doen. Het gevolg zou zijn een rho waarde van $1 - 1,27 = -0,27$. Een even sterk verband, zij het in omgekeerde richting.

Als we alleen nominale gegevens hebben, kunnen we uit een kruistabel al opmaken of er sprake is van een verband. In onderstaande tabel lijkt het er op dat mannen eerder voorstander zijn, vrouwen eerder tegenstander. Er is dus een verband tussen geslacht en opinie. In de inductieve statistiek wordt veel gebruik gemaakt van chi kwadraat. Deze geeft alleen aan of het verband significant is, niet de sterkte van het verband. De sterkte kan echter wel berekend worden met de van chi kwadraat

afgeleide maten phi en Cramer's V. In dit voorbeeld bedragen deze beide 0,36.

| | Man | Vrouw |
|--------|-------------|--------------|
| Ja | 200 (63,5%) | 100 (27,4%) |
| Nee | 115 (36,5%) | 265 (72,6%) |
| Totaal | 315 (100%) | 365 (100%) |

8.3 Causale uitspraken

We zagen dat er een redelijke correlatie bestaat tussen leenfrequentie en downloadfrequentie. Maar een correlatie is nog geen causale relatie. Het is natuurlijk niet zo dat men meer gaat downloaden omdat men veel leent, of omgekeerd. Hier is een derde achterliggende factor in het spel: studiemotivatie, hier gemeten/geoperationaliseerd aan de hand van het aantal studie-uren. Mogelijk spelen ook andere, niet gemeten zaken een rol op de achtergrond. Dat een correlatie nog geen causale relatie is blijkt ook uit een ander voorbeeld: er is een sterk verband tussen de mate waarin het asfalt smelt en de ijsconsumptie. De gemeenschappelijke oorzaak is uiteraard de hoge temperatuur.

Veel zinvoller lijkt de veronderstelling dat er een causale relatie bestaat tussen het volgen van een bibliotheekcursus enerzijds en uitleenfrequentie dan wel downloadfrequentie anderzijds. Misschien is het zo dat mensen na het volgen van een cursus eerder geneigd zijn documenten te lenen of te downloaden. Maar evengoed is het mogelijk dat juist de mensen die veel lenen of downloaden belangstelling hebben voor een bibliotheekcursus. De tijdsvolgorde is onbekend: eerst een cursus, daarna meer lenen of downloaden? Of eerst veel lenen of downloaden, daarna een cursus?

Uit deze twee voorbeelden blijkt dat je aan de hand van een vragenlijst nooit causale uitspraken kunt doen. Een andere methode, het experiment, leent zich daartoe veel beter. In een experiment kun je storende factoren zoveel mogelijk uitsluiten en heb je invloed op de tijdsvolgorde. Het meest haalbare bij een vragenlijst is het weg verklaren van mogelijke oorzaken. Het lijkt plausibel om een causaal verband te veronderstellen tussen studiemotivatie, gemeten in het aantal studie-uren enerzijds en leen- dan wel downloadfrequentie anderzijds. Maar misschien hangt het

aantal studie-uren sterk samen met de studierichting, en moeten we eerder denken aan een verband tussen studierichting en lenen/downloaden. Als we gegevens hebben over de studierichting kunnen we dat onderzoeken. Maar dan zijn nog weer andere storende factoren denkbaar, zoals leesattitude of beschikking over alternatieve voorzieningen. Die kunnen we onmogelijk allemaal in een vragenlijst opnemen. Ook dit illustreert weer het belang van een goed conceptueel model. Dat stelt ons in staat relevante vragen te stellen en, via correlatieonderzoek, causale verbanden niet zozeer aan te tonen als wel uit te sluiten.

9 Klanttevredenheidsonderzoek

Gebruikersonderzoek kan zich richten op gedrag, kennis, mening of tevredenheid van leden van de populatie. De laatste jaren is de term klanttevredenheidsonderzoek erg in zwang. Daarbij gaat het vooral om de mening of mate van tevredenheid van de klanten. Deze paragraaf biedt een overzicht van een aantal methoden.

9.1 Eendimensionale vragenlijst

Vaak bestaat een klanttevredenheidsvragenlijst uit stellingen. Respondenten kunnen via een schaalwaarde aangeven in hoeverre zij het daarmee eens zijn. Een ouder voorbeeld van Brooks, Revill en Shelton (1997) is nog steeds illustratief.

| | |
|--|----------------------|
| The library staff are knowledgeable | 1 2 3 4 5 6 7 |
| The library staff are unhelpful | 1 2 3 4 5 6 7 |
| Loan periods on key items are appropriate | 1 2 3 4 5 6 7 |
| Counter waiting times are acceptable | 1 2 3 4 5 6 7 |
| Up-to-date journals are readily available | 1 2 3 4 5 6 7 |
| The lighting is poor | 1 2 3 4 5 6 7 |
| It is safe to leave personal items unattended | 1 2 3 4 5 6 7 |
| The library contains sufficient items relevant to my studies | 1 2 3 4 5 6 7 |
| Too many items have restrictions on borrowing | 1 2 3 4 5 6 7 |
| Overall, the quality of the library is excellent | 1 2 3 4 5 6 7 |

De vragen zijn afgestemd op de gebruiker. Niet: ‘hoe goed vindt u onze collectie’, maar een stelling als ‘de bibliotheek bevat voldoende materiaal om in mijn studiebehoeften te voorzien’. Ook is er af en toe een negatieve bewering ingelast om een antwoordcadenans te voorkomen (‘De verlichting is slecht’). De laatste vraag naar een totaaloordeel is bedoeld om de validiteit van de vragenlijst te testen. Als het oordeel over de afzonderlijke onderdelen positiever of negatiever uitvalt dan het totaaloordeel, zijn kennelijk niet de juiste vragen gesteld.

Kiran en Diljit (2012) geven een voorbeeld van een meer eigentijdse vragenlijst. Deze is bedoeld om de ‘web-based library quality’ te meten. Enkele beweringen hieruit:

- The service provides trusted information compared to the Internet.
- The service provides access to a wide range of electronic resources in my subject area.
- The website is easy to use.
- There are enough ports for laptop use to access Web-based services.
- Online librarians understand my specific information needs.
- I am able to set up an alert for new materials in my discipline.
- The site allows me the convenience of sending a query/comment online.
- Using Web-based services, I can easily get what I am looking for most of the time.
- The Web-based services have innovative features that are interesting to use.

9.2 *Probleemidentificatie*

Een belangrijk doel van klanttevredenheidsonderzoek is om problemen te identificeren en vast te stellen op welke onderdelen verbetering noodzakelijk is. Er zijn andere opties dan schaalwaarden om dit te bereiken:

- Aankruisen van onderdelen die verbetering behoeven. Vraag bijvoorbeeld een algemeen oordeel over de studiezaal en leg de respondenten een lijst van details voor met het verzoek datgene aan te kruisen wat verbetering behoeft (aantal zitplaatsen, verlichting, meubilair, rust, temperatuur, aanwezigheid stopcontacten etc.).

- Stel (naast gesloten ook) open vragen, zoals:
 - Wat doen we goed?
 - Wat doen we niet goed?
 - Als we één onderdeel moeten verbeteren, wat zou dat dan zijn?
- Neem naast een vragenlijst ook interviews af. Deze kunnen ingaan op een 'critical incident': de ervaringen tijdens het laatste bezoek, de keer dat men ontevreden of juist tevreden was, of de ervaringen met de instelling tijdens het laatste project of de laatste studieopdracht.

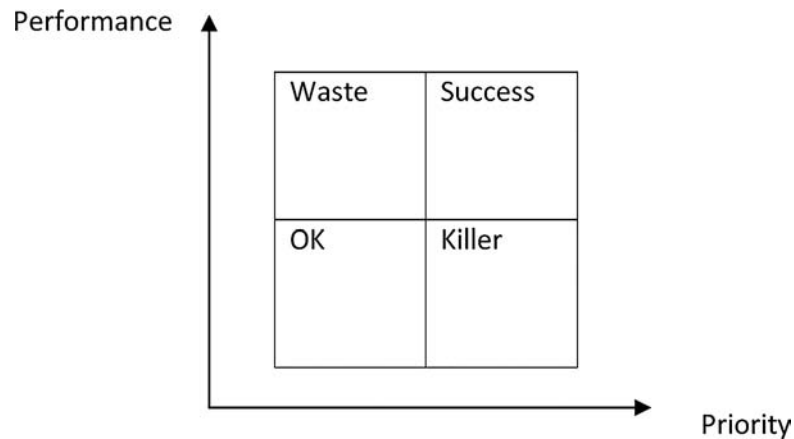
9.3 Tweedimensionale vragenlijst: tevredenheid en belang

De aanpak van Broadbent en Lofgren (1993) leent zich er ook goed toe om problemen te identificeren. Zij legden gebruikers een aantal aspecten voor met het verzoek om deze twee keer te beoordelen: niet alleen de mate van tevredenheid, maar ook het belang.

| | | |
|--|----------------------|--------------|
| A. In your opinion, what priority should ... give each of the following areas? | | |
| 1 low priority | 7 very high priority | I don't know |
| b. In your opinion, how well does ... perform in each of the following areas? | | |
| 1 very poorly | 7 excellently | I don't know |

Zij noemen hun aanpak dan ook Priority and Performance Evaluation (PAPE). De resultaten worden in een aansprekende figuur weergegeven. Deze bestaat uit vier kwadranten. In het kwadrant Killer bevinden zich die aspecten die men belangrijk vindt, maar waarover men niet tevreden is. Het kwadrant Success herbergt de kritische succesfactoren: belangrijke onderdelen waarop de instelling naar de mening van de respondenten ook goed functioneert. Onderdelen gelden als Waste als respondenten er wel tevreden over zijn, maar er weinig belang aan hechten. Indien daar veel kosten mee gemoeid zijn, kan men daar misschien op bezuinigen.

Voor het kwadrant OK geldt dat men niet tevreden is, maar er ook weinig belang aan hecht.



Meer recent volgden Pinto et al. (2010) deze aanpak. Zij vroegen naar het belang van en de tevredenheid over zaken als de gebruikersvriendelijkheid van de website, het aanbod van elektronische tijdschriften en de kwaliteit van digitale bibliotheekinstructie.

9.4 Tweedimensionale vragenlijst: ervaring versus verwachting

Tevredenheid kan worden opgevat als een vergelijking van ervaring en verwachting. Indien de feitelijk ervaren dienstverlening de verwachting overtreft, is men tevreden. Indien de feitelijk ervaren dienstverlening achter blijft bij de verwachting, is men ontevreden. Men kan de mate van tevredenheid dus meten door vragen te stellen over ervaring en verwachting, en daartussen de mate van overeenkomst dan wel discrepantie te berekenen. Het bekendste voorbeeld dat stoelt op dit uitgangspunt is LibQual+. Deze bevat 22 beweringen, zoals:

- Employees who understand the needs of their users.
- Modern equipment that lets me easily access needed information.
- A comfortable and inviting location.

De vragen moeten vanuit drie verschillende gezichtspunten beantwoord worden, op een negenpuntsschaal: het minimaal acceptabele niveau, het gewenste niveau, het daadwerkelijk ervaren niveau. Daaruit worden twee

scores afgeleid: de adequacy gap (het verschil tussen het daadwerkelijke ervaren en het minimaal acceptabele niveau) en de superiority gap (het verschil tussen het daadwerkelijk ervaren en het gewenste niveau). Er is een probleem indien de adequacy gap negatief is (de dienstverlening voldoet niet aan de meest elementaire verwachting). In de meeste gevallen is sprake van een superiority gap (alles kan altijd beter).

LibQual+ wordt veelvuldig toegepast door bibliotheken. Er is een hausse aan literatuur over dit instrument. Voorbij (2013) geeft elders in dit Handboek een beknopt overzicht.

9.5 *Causaal model*

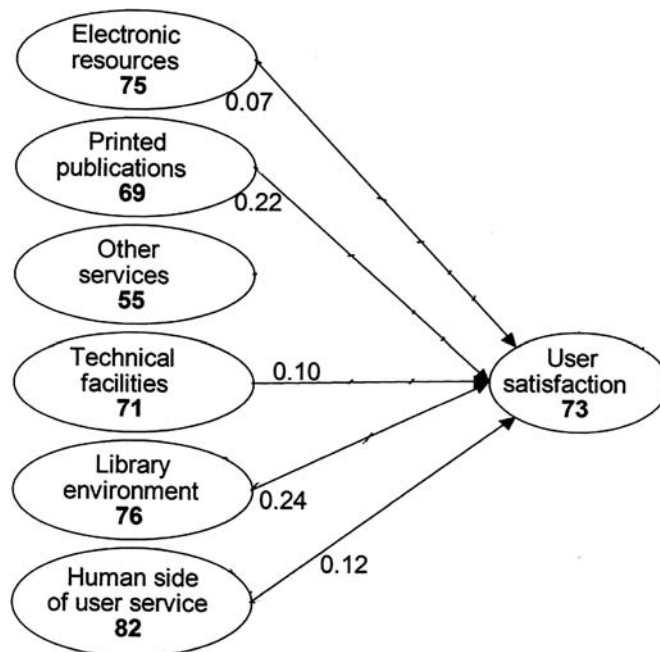
Het causale model van Martensen en Grønholdt (2003) heeft veel weg van het conceptuele model dat het uitgangspunt was van de vragenlijst naar het gebruik van online nieuwsbronnen (figuur 1). Tevredenheid wordt hier opgevat als een afhankelijke variabele, die beïnvloed wordt door een aantal onafhankelijke variabelen. De auteurs onderscheiden er een zestal. Aan elk van de variabelen is een beperkt aantal vragen gewijd (met mogelijkheden tot uitbreiding). De vragen over tevredenheid luiden:

- Considering all your experience of ..., how satisfied are you in general?
- To what degree do you consider that ... fulfills your expectations?
- Imagine a library which is perfect in all aspects. How close to this ideal do you consider to be?

Met deze methode krijgen we, net als bij de andere methoden, inzicht in de tevredenheid over diverse aspecten. Het bijzondere ligt erin dat we ook zien welk effect elk aspect heeft op de algemene tevredenheid. In figuur 6 zien we voor elk onderdeel een waardering op een schaal van 1 tot 100. De waarden op de lijnen geven de sterkte weer van de relatie tussen een onafhankelijke variabele en de tevredenheid (bij één variabele is geen waarde vermeld omdat deze te laag is). Deze 'impact scores' zijn gemeten aan de hand van regressieanalyse. In dit voorbeeld zien we onder meer dat het onderdeel gedrukte publicaties gewaardeerd wordt met een score van 69 en een vrij sterk effect heeft op de algemene tevredenheid (een impact score van 0,22). Op het eerste gezicht lijkt het merkwaardig dat een hoger gewaardeerd onderdeel als elektronische bronnen (75) een lagere impact score heeft (0,07). Een oorzaak kan zijn dat men wel tevreden is, maar weinig belang hecht aan de voorziening. Dit lijkt hier niet waarschijnlijk. Een andere oorzaak kan zijn dat we hier te maken

hebben met een 'dissatisfier'. Sommige diensten zijn zo basaal dat mensen kwaliteit voor kennisgeving aannemen. Dit geldt bijvoorbeeld voor een website zonder dode links. Dit is geen bron van tevredenheid, maar omgekeerd leidt een website met dode links wel tot ontevredenheid. Mogelijk werkt dat ook zo voor elektronische bronnen: een bibliotheek kan daar gewoon niet omheen en kan zich door big deals steeds minder onderscheiden van andere bibliotheken. De term dissatisfier impliceert dat je er eigenlijk alleen de mist mee kunt ingaan: een goede dienstverlening wordt niet beloond, een slechte wel afgestraft.

Figuur 6 is een vereenvoudigde weergave van de oorspronkelijke figuur. Martensen en Grønholdt onderscheiden niet één, maar drie afhankelijke variabelen: naast tevredenheid ook belang (value) en loyaliteit.



Figuur 6. Vereenvoudigde weergave van het model van Martensen en Grønholdt (2003).

10 Beperkingen

Gebruikersonderzoek is van groot belang. We verrichten onze activiteiten nu eenmaal voor de gebruiker en willen dan ook graag zijn stem horen. We moeten echter niet blind zijn voor de beperkingen. Voor een deel is de onderzoeker zelf verantwoordelijk voor het succes of gebrek aan succes van een gebruikersonderzoek. Zijn de vragen zodanig dat deze een goed inzicht geven in gedrag, motieven, behoeften, kennis of mening van de doelgroep? Is de steekproef voldoende groot en op aselectieve wijze samengesteld?

Voor een deel echter zijn we afhankelijk van de gebruikers.

1. *Lage en/of selectieve respons*

Door het gemak waarmee enquêtes tegenwoordig verspreid en verwerkt kunnen worden, ontvangen mensen meer uitnodigingen om vragenlijsten in te vullen dan hen lief is. Met als gevolg vaak een lage respons. Mogelijk ook is de respons selectief: juist personen met een uitgesproken mening zijn geneigd te reageren. Alle inspanningen die we hebben gedaan om een aselecte steekproef samen te stellen lopen gevaar om door de feitelijke respons teniet gedaan te worden. Het is maar zeer de vraag in hoeverre de resultaten van het onderzoek representatief zijn voor de gehele populatie.

2. *Incompetentie*

Het is niet moeilijk een oordeel te hebben over openingstijden, wel over bijvoorbeeld de kwaliteit van de collectie of het aanbod van digitale bronnen. Een bevinding van Falcone en Rivera (2005) is in dit opzicht illustratief. Uit hun gebruikersonderzoek kwam als grootste bron van ontevredenheid naar voren dat de bibliotheek niet zou beschikken over volledige sets van tijdschriften. Uit een usability study bleek echter dat zij moeite hadden om elektronische bronnen in de digitale bibliotheek te vinden. Uiteindelijk bleek het probleem dus eerder te maken te hebben met gebruikersvaardigheden of de systeeminterface dan de collectie; de onvrede over het tijdschriftenaanbod was onterecht. Dit illustreert dat het nog maar de vraag is in hoeverre de resultaten van gebruikersonderzoek betrouwbaar zijn.

3. *Conditionering*

Gebruikers zijn geconditioneerd, gewend aan de huidige voorzieningen en het huidige niveau van dienstverlening. Het is voor hen soms moeilijk zich voor te stellen dat het ook beter kan. Daarom is het maar de vraag hoeveel waarde we kunnen hechten aan gebruikers die zeggen tevreden te zijn. Een conservatieve houding is daarom op zijn plaats: ontevredenheid wijst sowieso op een probleem, tevredenheid kan mogelijk ten onrechte zijn (false positive). Instellingen moeten altijd zelf alert blijven op mogelijke problemen en verbeteringen en zich niet in slaap laten sussen door tevreden gebruikers.

Literatuur

- Broadbent, Marianne and Hans Lofgren. Information delivery: identifying priorities, performance and value. *Information Processing & Management* 29 (1993) 6, 683-701.
- Brooks, P., D. Reville and T. Shelton. The development of scale to measure the quality of an academic library from the perspective of its users. In J. Brockman (Ed.), *Quality Management and Benchmarking in the Information Sector, Results of Recent Research*. British Library Research and Innovation Report 47. London: British Library, 1997.
- Burkell, Jacquelyn. The dilemma of survey nonresponse. *Library & Information Science Research* 25 (2003) 3, 239-263.
- Falcone, Santa and Michael Rivera. Improving university library electronic services. *Performance Measurement and Metrics* 6 (2005) 2, 97-107.
- Harley, Diane and Jonathan Henke. Toward an effective understanding of website users. *D-Lib Magazine* 13 (2007) 3/4.
<http://www.dlib.org/dlib/march07/harley/03harley.html>
- Hider, Philip and Bob Pymm. Empirical research methods reported in high profile LIS journal literature. *Library & Information Science Research* 30 (2008) 2, 108-114.
- Julien, Heidi, Jen Pecoskie and Kathleen Reed. Trends in information behavior research, 1999-2008: A content analysis. *Library and Information Science Research* 33 (2011) 1, 19-24.
- Kim, Jihyun. Motivations of Faculty Self-archiving in Institutional Repositories. *Journal of Academic Librarianship* 37 (2011) 3, 246-254.
- King, Donald W. et al. Scholarly journal information-seeking and reading

- patterns of faculty at five US Universities. *Learned Publishing* 22 (2009) 2, 126-144.
- Kiran, K. and S. Diljit. Modeling web-based library service quality. *Library and Information Science Research* 34 (2012) 3, 184-196.
- Leefsma, Saskia. *Postuum beheer een verkennend onderzoek naar de ontsluiting van nagelaten collecties van kleine meesters*. Masterscriptie documentaire informatiewetenschap, Universiteit van Amsterdam, 2010.
<http://dare.uva.nl/cgi/arno/show.cgi?fid=170149>
- Martensen, Anne and Lars Grønholdt. Improving library users' perceived quality, satisfaction and loyalty: an integrated measurement and management system. *Journal of Academic Librarianship* 29 (2003) 3, 140-147.
- Mill, David H. Undergraduate information resource choices. *College & Research Libraries* 69 (2008) 4, 342-355.
<http://crl.acrl.org/content/69/4/342.short>
- Pinto, Maria, Viviana Fernández-Marcial, Carmen Gómez-Camarero. The impact of information behavior in academic library service quality: a case study of the science and technology areas in Spain. *Journal of Academic Librarianship* 36 (2010) 1, 70-78.
- Skov, Mette. Hobby-related information-seeking behavior of highly dedicated online museum visitors. *Information Research* 18 (2013) 4.
<http://www.informationr.net/ir/18-4/paper597.html#.VO81-xvmG8eo>
- Stieg Dalton, Margaret and Laurie Charnigo. Historians and Their Information Sources *College & Research Libraries* 65 (2004) 5, 400-425.
<http://crl.acrl.org/content/65/5/400.full.pdf>
- Tuamaala, Otto, Kalervo Järvelin and Pertti Vakkari. Evolution of library and information Science, 1965-2005: content analysis of journal articles. *Journal of the Association for Information Science and Technology* 65 (2014) 7, 1446-1462.
- Turcios, Mirna E., Naresh Kumar Agarwal and Linda Watkins. How much of library and information science literature qualifies as research? *Journal of Academic Librarianship* 40 (2014) 5, 473-479.
- Voorbij, Henk. Prestatie-indicatoren van bibliotheken. *Handboek Informatiewetenschap voor bibliotheek en archief*. (red) GM van Trier et al. Alphen aan de Rijn: Samsom, juli 2013, II 651 1- II 651 55.
www.iwabase.nl

Noten

1. <http://stats.oecd.org/glossary/detail.asp?ID=2620>
2. <http://www.cbsnews.com/news/forbid-or-allow>
3. <http://www.tns-nipo.com/ons-aanbod/marktonderzoek/panel>
4. <http://www.probiblio.nl/producten-en-diensten/marketing-en-communicatie/klant-kennis/biebpanel.html>